

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2003, 03:49:05 ; Search time 308 seconds

(without alignments)
551.370 Million cell updates/sec

Title: us-09-803-719-222

Perfect score: 378
Sequence: 1 tacgcgtccttaagacgact.....taccatgtgagaacatgcg 378

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 396772 seqs, 224632407 residues

Word size : 0

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Published Applications_NA.*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	5.0	527	10 US-09-864-761-7775	Sequence 7775, App
2	19	5.0	709	9 US-09-910-943-287	Sequence 287, App
3	18	4.8	26555	10 US-09-860-670-161	Sequence 161, App
4	17	4.5	309	9 US-09-796-692-35	Sequence 35, Appl
5	17	4.5	309	9 US-09-796-692-3537	Sequence 3537, App
6	17	4.5	397	10 US-09-878-574-3537	Sequence 843, App
7	17	4.5	627	9 US-10-040-739-843	Sequence 1704, App
8	17	4.5	837	9 US-09-938-842A-1704	Sequence 1051, App
9	17	4.5	1005	10 US-09-867-550-1051	Sequence 31616, A
10	17	4.5	1041	10 US-09-864-761-31616	Sequence 3427, App
11	17	4.5	1359	9 US-09-938-842A-1704	Sequence 13, Appl
12	17	4.5	1666	10 US-09-739-254-13	Sequence 11, Appl
13	17	4.5	1666	10 US-09-904-615-13	Sequence 12, Appl
14	17	4.5	1966	10 US-09-864-761-15087	Sequence 15087, A
15	17	4.5	6560	9 US-09-808-602-76	Sequence 76, Appl
16	17	4.5	9729	9 US-09-808-602-76	Sequence 76, Appl
17	17	4.5	9826	9 US-09-808-602-76	Sequence 7, Appl
18	17	4.5	40392	10 US-09-954-456-44	Sequence 44, Appl
19	17	4.5	40392	10 US-09-954-456-687	Sequence 687, App

C 20	17	4.5	659158	9 US-09-771-208-20	Sequence 20, Appl
21	16	4.2	191	10 US-09-878-574-10027	Sequence 10027, A
22	16	4.2	261	10 US-09-878-574-12914	Sequence 12914, A
23	16	4.2	275	10 US-09-923-876-513	Sequence 513, App
C 24	16	4.2	403	9 US-10-046-935-15	Sequence 15, Appl
C 25	16	4.2	403	9 US-09-878-178-15	Sequence 15, Appl
C 26	16	4.2	443	10 US-09-833-381-571	Sequence 671, App
C 27	16	4.2	465	10 US-09-833-381-1080	Sequence 1080, App
C 28	16	4.2	507	10 US-09-864-761-7849	Sequence 7849, App
C 29	16	4.2	551	10 US-09-864-761-11919	Sequence 11919, A
C 30	16	4.2	583	10 US-09-864-761-14818	Sequence 14818, A
C 31	16	4.2	636	9 US-09-938-842A-1	Sequence 1, Appl1
C 32	16	4.2	869	10 US-09-770-445-565	Sequence 565, App
C 33	16	4.2	906	10 US-09-925-397-302	Sequence 302, App
C 34	16	4.2	936	10 US-09-886-055-354	Sequence 354, App
C 35	16	4.2	1006	9 US-09-964-899-22	Sequence 22, Appl
C 36	16	4.2	1212	9 US-09-938-842A-3713	Sequence 733, App
C 37	16	4.2	1650	10 US-09-880-107-2199	Sequence 2199, App
C 38	16	4.2	1758	10 US-09-731-872-72	Sequence 72, Appl1
C 39	16	4.2	1768	10 US-09-731-872-7	Sequence 107, App
C 40	16	4.2	1846	9 US-10-098-841-107	Sequence 7, Appl1
C 41	16	4.2	1992	9 US-10-080-960-15	Sequence 15, Appl
C 42	16	4.2	2000	9 US-09-938-842A-3712	Sequence 3712, App
C 43	16	4.2	2003	10 US-09-887-576-306	Sequence 306, App
C 44	16	4.2	2004	10 US-09-887-576-253	Sequence 253, App
C 45	16	4.2	2481	10 US-09-852-909-3	Sequence 3, Appl1

ALIGNMENTS

```
RESULT 1
US-09-864-761-7775
Sequence 7775, Application US/09864761
Patent No. US2002048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
```

RESULT 3
US-09-860-670-161
; Sequence 161, Application US/09860670
; Patent No. US20020165137A1
; GENERAL INFORMATION:
; APPLICANT: Ruden et al.

```

;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(309)
;   OTHER INFORMATION: n = A,T,C or G
US-09-796-692-35

```

Query Match:	4.58;	Score 17;	DB 9;	Length 309;
Best Local Similarity	100.08;	Pred. No. 12;		

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 94 AATGTGTCTGGAGC 110
|||||
Db 68 AATGTGTCTGGAGC 52

RESULT 5
US-09-796-692-3533/C
; Sequence 3533, Application US/09796692
; Publication No. US20020198562A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 3533
; LENGTH: 309
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (11)
; OTHER INFORMATION: n-A,T,C or G
; NAME/KEY: unsure
; LOCATION: (229)
; OTHER INFORMATION: n-A,T,C or G
; NAME/KEY: unsure
; LOCATION: (272)
; OTHER INFORMATION: n-A,T,C or G
; NAME/KEY: unsure
; LOCATION: (277)
; OTHER INFORMATION: n-A,T,C or G
; NAME/KEY: unsure
; LOCATION: (289)
; OTHER INFORMATION: n-A,T,C or G
; NAME/KEY: unsure
; LOCATION: (296)
; OTHER INFORMATION: n-A,T,C or G
US-09-796-692-3533

Query Match 4.5%; Score 17; DB 9; Length 309;
Best Local Similarity 100.0%; Pred. No. 12;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 94 AATGTGTCTGGAGC 110
|||||
Db 68 AATGTGTCTGGAGC 52

RESULT 6
US-09-878-574-3537
; Sequence 3537, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 3537
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(397)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3028-009-Q1-B1-B5
US-09-878-574-3537

Query Match 4.5%; Score 17; DB 10; Length 397;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 103 TCTGAGCTGAGGATG 119
|||||
Db 85 TCTGAGCTGAGGATG 101

RESULT 7
US-10-040-739-843/C
; Sequence 843, Application US/10040739
; Patent No. US20020173635A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallee, Edward
; APPLICANT: Racine, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
; NUMBER OF SEQUENCES: 1519
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/040,739
; FILING DATE: 07-Jan-2002
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/036,520
FILING DATE: 03-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 843:
SEQUENCE CHARACTERISTICS:
LENGTH: 627 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 843:
US-10-040-739-843

Query Match 4.5%: Score 17; DB 9; Length 627;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 269 GAGATGCCAATCCATGCG 285
|||||
DB 610 GAGATGCCAATCCATGCG 594

RESULT 8
US-09-938-842A-1704
Sequence 1704, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kieps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1704
LENGTH: 837
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1704

Query Match 4.5%: Score 17; DB 9; Length 837;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 229 TGAAGAAGACCTTGG 245
|||||
DB 276 TGAAGAAGACCTTGG 292

RESULT 9
US-09-867-550-1051/c
Sequence 1051, Application US/09867550
Patent No. US20020082206A1
GENERAL INFORMATION:
APPLICANT: Leach, Martin D.
APPLICANT: Mehraban, Fued,
APPLICANT: Conley, Pamela
APPLICANT: Law, Debbie

APPLICANT: Topper, James
TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cell
TITLE OF INVENTION: Thereby
FILE REFERENCE: 21402-013 (Cura-313)
CURRENT APPLICATION NUMBER: US/09/867,550
CURRENT FILING DATE: 2001-09-20
PRIOR APPLICATION NUMBER: USSN 60/208,427
PRIOR FILING DATE: 2000-05-30
NUMBER OF SEQ ID NOS: 2125
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1051
LENGTH: 1005
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)..(2)
OTHER INFORMATION: wherein any n is one of a or t or g or c
US-09-867-550-1051

Query Match 4.5%: Score 17; DB 10; Length 1005;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 271 GATGCCAATCCATGCGA 287
|||||
DB 600 GATGCCAATCCATGCGA 584

RESULT 10
US-09-864-761-31616/c
Sequence 31616, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30

```
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31616
; LENGTH: 1041
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004526.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9
; OTHER INFORMATION: SWISSPROT HIT: P34910, EVALU0.00e+00
; OTHER INFORMATION: NT HIT: 911426548, EVALU0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AA805896.1, EVALU0.00e+00
US-09-864-761-31616
```

```
Query Match 4.5%; Score 17; DB 10; Length 1041;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 301 CTATGTTGGTACTATA 317
Db 467 CTATGTTGGTACTATA 451
```

```
RESULT 11
US-09-938-842A-3427/C
; Sequence 3427, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: S000000-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3427
; LENGTH: 1359
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3427
```

```
Query Match 4.5%; Score 17; DB 9; Length 1359;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 76 GAGTCTCTCAACGCT 92
Db 531 GAGTCTCTCAACGCT 515
```

```
RESULT 12
US-09-739-254-13/C
; Sequence 13, Application US/09739254
; Patent No. US20010021700A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032P1
; CURRENT APPLICATION NUMBER: US/09/739,254
; CURRENT FILING DATE: 2000-12-19
; EARLIER APPLICATION NUMBER: 09/511,554
; EARLIER FILING DATE: 2000-02-23
; EARLIER APPLICATION NUMBER: PCT/US99/19330
; EARLIER FILING DATE: 1999-08-24
; EARLIER APPLICATION NUMBER: 60/097,917
; EARLIER FILING DATE: 1998-08-25
; EARLIER APPLICATION NUMBER: 60/098,634
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 13
; LENGTH: 1666
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-739-254-13
```

```
Query Match 4.5%; Score 17; DB 10; Length 1666;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 271 GATGCCATCATCGAA 287
Db 740 GATGCCATCATCGAA 724
```

```
RESULT 13
US-09-904-615-13/C
; Sequence 13, Application US/09904615
; Patent No. US20020026040A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032P1
; CURRENT APPLICATION NUMBER: US/09/904,615
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/511,554
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 13
; LENGTH: 1666
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-904-615-13
```

```
Query Match 4.5%; Score 17; DB 10; Length 1666;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 271 GATGCCATCATCGAA 287
Db 740 GATGCCATCATCGAA 724
```

```
RESULT 14
US-09-864-761-15087/C
; Sequence 15087, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
```

;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Aecm1ca-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 15087
;; LENGTH: 1966
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC004526.1
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9
US-09-864-761-15087
Query Match 4.5%; Score 17; DB 10; Length 1966;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 301 CTATGTTGTAAGTATA 317
DB 810 CTATGTTGTAAGTATA 794

RESULT 15
US-09-808-602-76/c
; Sequence 76, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A

;; APPLICANT: Herriman, John L
;; APPLICANT: Majumder, Kumud
;; APPLICANT: Mishra, Vishnu
;; APPLICANT: Mezes, Peter S
;; APPLICANT: MacDougall, John
;; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nucleic Acids Encoding Sar
;; FILE REFERENCE: 15966-697 CIP
;; CURRENT APPLICATION NUMBER: US/09/808,602
;; CURRENT FILING DATE: 2001-03-14
;; PRIOR APPLICATION NUMBER: 09/800,198
;; PRIOR FILING DATE: 2001-03-05
;; PRIOR APPLICATION NUMBER: 60/186,596
;; PRIOR FILING DATE: 2000-03-03
;; NUMBER OF SEQ ID NOS: 114
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 76
;; LENGTH: 6560
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-808-602-76
Query Match 4.5%; Score 17; DB 9; Length 6560;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 185 CTGCCAAGTACTACTG 201
DB 767 CTGCCAAGTACTACTG 751
Search completed: January 31, 2003, 05:19:41
Job time : 345 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2003, 01:19:33 : Search time 53 Seconds
(without alignments)
2187.242 Million cell updates/sec

Title: US-09-803-719-222

Perfect score: 378
Sequence: 1 tacgctgcttaagagacgact.....taccatgagagacatgacg 378

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size: 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Issued_Patents_NA:*

- 1: /cgn2_6/prodate/1/lna/5A.COMB.seq:*
- 2: /cgn2_6/prodate/1/lna/5B.COMB.seq:*
- 3: /cgn2_6/prodate/1/lna/6A.COMB.seq:*
- 4: /cgn2_6/prodate/1/lna/6B.COMB.seq:*
- 5: /cgn2_6/prodate/1/lna/PCTUS.COMB.seq:*
- 6: /cgn2_6/prodate/1/lna/Backlist1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	4.5	1080	4 US-09-134-001C-1264	Sequence 1264, Appl
2	16	4.2	237	1 US-08-480-910-6	Sequence 6, Appl
3	16	4.2	1146	4 US-09-257-580-1	Sequence 1, Appl
4	16	4.2	1146	4 US-09-257-580-4	Sequence 1, Appl
5	16	4.2	1285	1 US-08-480-910-1	Sequence 1, Appl
6	16	4.2	1285	5 PCT-US95-00052-1	Sequence 17, Appl
7	16	4.2	1314	4 US-09-500-569-17	Sequence 9, Appl
8	16	4.2	1873	4 US-09-461-474-9	Sequence 11, Appl
9	16	4.2	4413	4 US-09-221-017B-811	Sequence 11, Appl
10	16	4.2	18596	2 US-08-735-609-4	Sequence 4, Appl
11	16	4.2	34303	2 US-08-735-609-4	Sequence 4, Appl
12	16	4.2	34303	3 US-09-315-372-4	Sequence 4, Appl
13	16	4.2	34303	3 US-09-244-752-4	Sequence 4, Appl
14	16	4.2	34303	3 US-09-245-497-4	Sequence 4, Appl
15	16	4.2	34303	4 US-09-562-919-4	Sequence 6, Appl
16	16	4.2	34303	4 US-08-374-483-6	Sequence 3, Appl
17	16	4.2	34382	2 US-08-973-334-5	Sequence 3, Appl
18	16	4.2	35408	4 US-09-563-869A-3	Sequence 3, Appl
19	16	4.2	35408	4 US-08-549-489-3	Sequence 1, Appl
20	16	4.2	35935	2 US-08-735-609-1	Sequence 1, Appl
21	16	4.2	35935	2 US-08-735-609-1	Sequence 1, Appl
22	16	4.2	35935	3 US-08-379-452-43	Sequence 43, Appl
23	16	4.2	35935	3 US-09-315-372-1	Sequence 1, Appl
24	16	4.2	35935	3 US-09-244-752-1	Sequence 1, Appl
25	16	4.2	35935	3 US-09-245-497-1	Sequence 1, Appl
26	16	4.2	35935	3 US-09-245-497-1	Sequence 1, Appl
27	16	4.2	35935	3 US-09-409-670-43	Sequence 43, Appl

C 28	16	4.2	35935	4 US-09-562-919-1	Sequence 1, Appl
29	16	4.2	48974	4 US-08-920-422-17	Sequence 17, Appl
30	16	4.2	246240	2 US-08-724-394A-20	Sequence 20, Appl
31	16	4.2	246240	2 US-08-724-394A-21	Sequence 21, Appl
32	16	4.2	246240	2 US-08-724-394A-22	Sequence 22, Appl
33	15	4.0	22	2 US-08-734-941-2	Sequence 2, Appl
34	15	4.0	32	4 US-08-922-992A-29	Sequence 29, Appl
35	15	4.0	32	4 US-09-390-867A-25	Sequence 25, Appl
36	15	4.0	49	4 US-09-548-260-23	Sequence 23, Appl
37	15	4.0	81	4 US-08-895-590-23	Sequence 23, Appl
38	15	4.0	89	4 US-09-390-867A-23	Sequence 23, Appl
39	15	4.0	89	4 US-09-548-260-23	Sequence 23, Appl
40	15	4.0	280	3 US-09-026-343-31	Sequence 31, Appl
41	15	4.0	280	4 US-09-362-871-31	Sequence 31, Appl
42	15	4.0	344	4 US-08-644-426-3	Sequence 3, Appl
43	15	4.0	444	4 US-07-925-978B-1	Sequence 1, Appl
44	15	4.0	451	4 US-09-385-987-55	Sequence 55, Appl
45	15	4.0	540	3 US-09-042-771-3	Sequence 3, Appl

ALIGNMENTS

```

RESULT 1
US-09-134-001C-1264
; Sequence 1264, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1264
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1264

Query Match      4.5%  Score 17: DB 4: Length 1080;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 309 GTAGCTATACAGAGT 325
    |||||||||||
Db 889 GTAGCTATACAGAGT 905

RESULT 2
US-08-480-910-6
; Sequence 6, Application US/08480910
; Patent No. 5693530
; GENERAL INFORMATION:
; APPLICANT: Karel A. Schat, Kazuhiko Ohashi, and Priscilla
; TITLE OF INVENTION: A Marek's Disease Virus Nucleotide
; TITLE OF INVENTION: Sequence and Methods of Use
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
; STREET: 1800 One Met Plaza
; CITY: Buffalo
; STATE: New York
; COUNTRY: United States
; ZIP: 14203-2391
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Diskette, 3.5 inch, 1.44mb storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: MS-DOS/ Microsoft Windows
 SOFTWARE: Wordperfect for Windows 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/480,910
 FILING DATE: 07 June 1995
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: U.S. Serial No. 5693530 08/180,051
 FILING DATE: 11 January 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Nelson, M. Bud
 REGISTRATION NUMBER: 35,300
 REFERENCE/DOCKET NUMBER: 18617,0005
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (716) 849-0349
 TELEFAX: (716) 849-0349
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 237 nucleotides
 TYPE: nucleic acid
 STRANDEDNESS: single-stranded
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 US-08-480-910-6

Query Match 4.2%; Score 16; DB 1; Length 237;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 AGAACAATATGAGAAA 64
 |||||||
 DB 34 AGAACAATATGAGAAA 49

RESULT 3
 US-09-257-580-1
 Sequence 1, Application US/09257580
 Patent No. 6307036
 GENERAL INFORMATION:
 APPLICANT: Yorkshire Cancer Research
 TITLE OF INVENTION: Tumour Suppressor Gene
 FILE REFERENCE: Canine p53
 CURRENT APPLICATION NUMBER: US/09/257,580
 PRIOR FILING DATE: 1999-02-25
 PRIOR APPLICATION NUMBER: 9804178.3
 PRIOR FILING DATE: 1998-02-28
 NUMBER OF SEQ ID NOS: 11
 SOFTWARE: Patentln Ver. 2.1
 SEQ ID NO 1
 LENGTH: 1146
 TYPE: DNA
 ORGANISM: Canis
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)..(1146)
 US-09-257-580-1

Query Match 4.2%; Score 16; DB 4; Length 1146;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 TGGAGCTGAAGATGC 120
 |||||||
 DB 1007 TGGAGCTGAAGATGC 1022

RESULT 4
 US-09-257-580-4
 Sequence 4, Application US/09257580
 Patent No. 6307036
 GENERAL INFORMATION:
 APPLICANT: Yorkshire Cancer Research

TITLE OF INVENTION: Tumour Suppressor Gene
 FILE REFERENCE: Canine p53
 CURRENT APPLICATION NUMBER: US/09/257,580
 CURRENT FILING DATE: 1999-02-25
 PRIOR APPLICATION NUMBER: 9804178.3
 PRIOR FILING DATE: 1998-02-28
 NUMBER OF SEQ ID NOS: 11
 SOFTWARE: Patentln Ver. 2.1
 SEQ ID NO 4
 LENGTH: 1146
 TYPE: RNA
 ORGANISM: Canis
 US-09-257-580-4

Query Match 4.2%; Score 16; DB 4; Length 1146;
 Best Local Similarity 81.2%; Pred. No. 44;
 Matches 13; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 105 TGGAGCTGAAGATGC 120
 :|||||:|||||:11
 DB 1007 TGGAGCTGAAGATGC 1022

RESULT 5
 US-08-480-910-1
 Sequence 1, Application US/08480910
 Patent No. 5693530
 GENERAL INFORMATION:
 APPLICANT: Karel A. Schat, Kazuhiko Ohashi, and Priscilla
 APPLICANT: H. O'Connell
 TITLE OF INVENTION: A Marek's Disease Virus Nucleotide
 TITLE OF INVENTION: Sequence and Methods of Use
 Patent No. 5693530
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
 STREET: 1800 One Met Plaza
 CITY: Buffalo
 STATE: New York
 COUNTRY: United States
 ZIP: 14203-2391
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 inch, 1.44mb storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: MS-DOS/ Microsoft Windows
 SOFTWARE: Wordperfect for Windows 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/480,910
 FILING DATE: 07 June 1995
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: U.S. Serial No. 5693530 08/180,051
 FILING DATE: 11 January 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Nelson, M. Bud
 REGISTRATION NUMBER: 35,300
 REFERENCE/DOCKET NUMBER: 18617,0005
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (716) 849-0349
 TELEFAX: (716) 849-0349
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1285 nucleotides
 TYPE: nucleic acid
 STRANDEDNESS: single-stranded
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: yes
 IMMEDIATE SOURCE:
 LIBRARY: cDNA
 CLONE: clone 11
 ORIGINAL SOURCE:
 ORGANISM: Marek's Disease Virus
 STRAIN: MDC-CU41

CELL TYPE: VIRUS
FEATURE: L1 open reading frame, 725-1045
IDENTIFICATION METHOD: by experiment
OTHER INFORMATION:
US-08-480-910-1

Query Match 4.2%; Score 16; DB 1; Length 1285;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 AGGACATATGAAAA 64
|||||
Db 761 AGGACATATGAAAA 776

RESULT 6
PCT-US95-00052-1

Sequence 1, Application PC/RTUS9500052
GENERAL INFORMATION:

APPLICANT: Karel A. Schat, Kazuhiko Ohashi, and

TITLE OF INVENTION: Control Of Marek's Disease By The

TITLE OF INVENTION: Inhibition Of Latency And Tumor Cell Development

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hodgson, Russ, Andrews, Woods &

ADDRESSEE: Goodyear

STREET: 1800 One Mt Plaza

CITY: Buffalo

STATE: New York

COUNTRY: United States

ZIP: 14203-2391

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb storage

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1

SOFTWARE: Wordperfect for Windows 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/00052

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: U.S. Serial No. 08/180,051

FILING DATE: January 11, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Nelson, M. Bud

REGISTRATION NUMBER: 35,300

REFERENCE/DOCKET NUMBER: 18617.0005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (716) 856-4000

TELEFAX: (716) 849-0349

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1285 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single-stranded

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: yes

IMMEDIATE SOURCE:

LIBRARY: cDNA

CLONE: clone L1

ORIGINAL SOURCE:

ORGANISM: Marek's Disease Virus

STRAIN: MDC-CU41

CELL TYPE: virus

FEATURE:

LOCATION: L1 open reading frame, 725-1045

IDENTIFICATION METHOD: by experiment

OTHER INFORMATION:

PCT-US95-00052-1

Query Match 4.2%; Score 16; DB 5; Length 1285;

Best Local Similarity 100.0%; Pred. No. 44;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 AGGACATATGAAAA 64
|||||
Db 761 AGGACATATGAAAA 776

RESULT 7
US-09-500-569-17/c

Sequence 17, Application US/09500569

PATENT No. 6329204

GENERAL INFORMATION:

APPLICANT: Cahoon, Rebecca E.

APPLICANT: Rafalski, Antoni

APPLICANT: Shen, Jennie

TITLE OF INVENTION: Plant Caffeic acid 3-O-Methyltransferase Homologs

FILE REFERENCE: B01327 US NA

CURRENT APPLICATION NUMBER: US/09/500,569

CURRENT FILING DATE: 2000-02-09

EARLIER APPLICATION NUMBER: 60/119,587

EARLIER FILING DATE: 1999-February-10

NUMBER OF SEQ ID NOS: 28

SOFTWARE: Microsoft Office 97

SEQ ID NO 17

LENGTH: 1314

TYPE: DNA

ORGANISM: Glycine max

FEATURE:

NAME/KEY: unsure

LOCATION: (472)

FEATURE:

NAME/KEY: unsure

LOCATION: (1156)

FEATURE:

NAME/KEY: unsure

LOCATION: (1180)

FEATURE:

NAME/KEY: (1262)

FEATURE:

NAME/KEY: unsure

LOCATION: (1302)

US-09-500-569-17

Query Match 4.2%; Score 16; DB 4; Length 1314;

Best Local Similarity 100.0%; Pred. No. 44;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 67 CCCAGCGGAGCTCTC 82
|||||

Db 1247 CCCAGCGGAGCTCTC 1232

RESULT 8

US-09-461-474-9/c

Sequence 9, Application US/09461474

PATENT No. 6278042

GENERAL INFORMATION:

APPLICANT: Allen, Steve

APPLICANT: Rafalski, Antoni

APPLICANT: Sakai, Hajime

TITLE OF INVENTION: Plant Metal Transporters

FILE REFERENCE: B01303 US NA

CURRENT APPLICATION NUMBER: US/09/461,474

CURRENT FILING DATE: 1999-12-14

EARLIER APPLICATION NUMBER: 60/112,562

EARLIER FILING DATE: 1998-12-16

NUMBER OF SEQ ID NOS: 17

SOFTWARE: Microsoft Office 97

SEQ ID NO 9

LENGTH: 1873

TYPE: DNA

ORGANISM: Oryza sativa
US-09-461-474-9

Query Match 4.2%; Score 16; DB 4; Length 1873;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 258 GGATGCTGATGAGAT 273
DB 885 GGATGCTGATGAGAT 870

RESULT 9
US-09-221-017B-811/C
Sequence 811, Application US/09221017B

PATENT No. 6444799
GENERAL INFORMATION:
APPLICANT: ROSS, BRUCE C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CA.
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221.017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, GLADYS H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 811:
SEQUENCE CHARACTERISTICS:
LENGTH: 4413 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...4413
US-09-221-017B-811
Query Match 4.2%; Score 16; DB 4; Length 4413;

Best Local Similarity 100.0%; Pred. No. 41;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 229 TGAAGAAGACCTTG 244
DB 845 TGAAGAAGACCTTG 830

RESULT 10
US-09-318-448-11
Sequence 11, Application US/09318448

PATENT No. 6210950
GENERAL INFORMATION:
APPLICANT: JOHNSON, WILLIAM G.
TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
FILE REFERENCE: 601-1-057
CURRENT APPLICATION NUMBER: US/09/318.448
CURRENT FILING DATE: 1999-05-25
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 18596
TYPE: DNA
ORGANISM: Homo sapiens
US-09-318-448-11

Query Match 4.2%; Score 16; DB 4; Length 18596;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 259 GATGCTGATGAGATG 274
DB 10134 GATGCTGATGAGATG 10149

RESULT 11
US-08-735-609-4/C
Sequence 4, Application US/08735609

PATENT No. 595360
GENERAL INFORMATION:
APPLICANT: Chamberlain, Jeffrey S.
APPLICANT: Amalfitano, Andrea
APPLICANT: Hauser, Michael A.
APPLICANT: Kumar-Singh, Rajendra
APPLICANT: Hartigan-O'Connor, Dennis J.
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/735.609
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 34303 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-08-735-609-4

Query Match 4.2%; Score 16; DB 2; Length 34303;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 212 CAACATGCTTGACTG 227
DB 27052 CAACATGCTTGACTG 27037

RESULT 12
US-08-735-609-4/c
Sequence 4, Application US/08735609
Patent No. 5994132
GENERAL INFORMATION:
APPLICANT: Chamberlain, Jeffrey S.
Amalfitano, Andrea
Hauser, Michael A.
Kumar-Singh, Rajendra
Hartigan-O'Connor, Dennis J.

TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/735,609
FILING DATE: 23-Oct-1996
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 34303 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-08-735-609-4

Query Match 4.2%; Score 16; DB 2; Length 34303;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 212 CAACATGCTTGACTG 227
DB 27052 CAACATGCTTGACTG 27037

RESULT 13
US-09-315-372-4/c
Sequence 4, Application US/09315372
Patent No. 6057158
GENERAL INFORMATION:
APPLICANT: Chamberlain, Jeffrey S.
Amalfitano, Andrea
Hauser, Michael A.
Kumar-Singh, Rajendra
Hartigan-O'Connor, Dennis J.

TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/315,372
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/735,609
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 34303 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-315-372-4

Query Match 4.2%; Score 16; DB 3; Length 34303;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 212 CAACATGCTTGACTG 227
DB 27052 CAACATGCTTGACTG 27037

RESULT 14
US-09-244-752-4/c
Sequence 4, Application US/09244752
Patent No. 6063622
GENERAL INFORMATION:
APPLICANT: Chamberlain, Jeffrey S.
Amalfitano, Andrea
Hauser, Michael A.
Kumar-Singh, Rajendra
Hartigan-O'Connor, Dennis J.

TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco

```

STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,752
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/735,609
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 34303 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-244-752-4

```

```

Query Match      4.2%; Score 16; DB 3; Length 34303;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 212 CAACATGCTTTGACTG 227
Db 27052 CAACATGCTTTGACTG 27037

```

```

RESULT 15
US-09-245-497-4/C
Sequence 4, Application US/09245497
Patent No. 6083750
GENERAL INFORMATION:
APPLICANT: Chamberlain, Jeffrey S.
APPLICANT: Amalfitano, Andrea
APPLICANT: Hauser, Michael A.
APPLICANT: Kumar-Singh, Rajendra
APPLICANT: Hartigan-O'Connor, Dennis J.
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medien & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/245,497
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/735,609
APPLICATION NUMBER: <B> FILING DATE:
ATTORNEY/AGENT INFORMATION:

```

```

NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 34303 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-09-245-497-4

```

```

Query Match      4.2%; Score 16; DB 3; Length 34303;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 212 CAACATGCTTTGACTG 227
Db 27052 CAACATGCTTTGACTG 27037

```

```

Search completed: January 31, 2003, 03:51:24
Job time : 133 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2003, 03:22:50 ; Search time 2218 Seconds

(without alignments)
2760.096 Million cell updates/sec

Title: US-09-803-719-222

Perfect score: 378

Sequence: 1 tacgctgtcttaagacgact.....taccattggagaacatgcy 378

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	90	23.8	451	14	BM987874
C 2	49	13.0	386	9	AA743908
C 3	49	13.0	439	9	AA743908
C 4	45	11.9	806	12	AI459918
C 5	34	9.0	349	17	AI459918
C 6	30	7.9	237	17	AI459918

C 7	21	5.6	357	12	BC210731	BC210731	RSR30277
C 8	21	5.6	440	14	BO597872	BO597872	MI-P-A2-a
C 9	21	5.6	526	12	BC223756	BC223756	1M00018F1
C 10	21	5.6	541	10	AV605455	AV605455	AV605455
C 11	21	5.6	545	10	AV605117	AV605117	AV605117
C 12	21	5.6	547	12	BF713089	BF713089	MI-P-H3-a
C 13	21	5.6	569	10	AM464207	AM464207	BP230015A
C 14	21	5.6	589	17	AZ364217	AZ364217	1M010P18
C 15	21	5.6	611	14	BO600629	BO600629	MI-P-E7-a
C 16	21	5.6	628	14	BO599839	BO599839	MI-P-E7-a
C 17	21	5.6	635	14	BO597758	BO597758	MI-P-E7-a
C 18	21	5.6	666	10	AV610564	AV610564	AV610564
C 19	21	5.6	680	13	B1186343	B1186343	UNL-P-FN-
C 20	21	5.6	692	14	BO604599	BO604599	MI-P-E7-a
C 21	21	5.6	785	13	B1184120	B1184120	UNL-P-FN-
C 22	20	5.3	379	10	BE769032	BE769032	PM4-F7002
C 23	20	5.3	391	9	AL672659	AL672659	MI-P-E7-a
C 24	20	5.3	391	9	AL672659	AL672659	MI-P-E7-a
C 25	20	5.3	440	17	AQ204259	AQ204259	HS-3116-B
C 26	20	5.3	507	9	AL798424	AL798424	AL798424
C 27	20	5.3	542	9	AL638747	AL638747	AL638747
C 28	20	5.3	640	13	BO394175	BO394175	BO394175
C 29	20	5.3	769	17	AO916538	AO916538	nbeb0065M
C 30	20	5.3	812	17	BH703215	BH703215	BOA0707F
C 31	20	5.3	847	14	BO885975	BO885975	ACBENCOURT
C 32	19	5.0	310	12	BF760676	BF760676	CAI-CT063
C 33	19	5.0	334	13	BM443382	BM443382	EBR002-SQ
C 34	19	5.0	347	17	AZ652179	AZ652179	IM0525J15
C 35	19	5.0	356	14	C28465	C28465	C28465
C 36	19	5.0	363	12	BF362917	BF362917	PM4-NN009
C 37	19	5.0	369	17	AO524481	AO524481	HS-5214-B
C 38	19	5.0	436	14	BO762497	BO762497	EBR002-SQ
C 39	19	5.0	506	17	AO441231	AO441231	HS-5129-B
C 40	19	5.0	523	9	AL697366	AL697366	AL697366
C 41	19	5.0	534	17	AF175356	AF175356	AF175356
C 42	19	5.0	574	13	BM657418	BM657418	170006874
C 43	19	5.0	578	14	BO579887	BO579887	ACB06-1m7
C 44	19	5.0	616	12	BG409744	BG409744	510-4-A1
C 45	19	5.0	631	13	BM635028	BM635028	170006875

ALIGNMENTS

RESULT 1
BM987874 451 bp mRNA linear EST 17-JUN-2002
DEFINITION
IMAGE:5860149 3', mRNA sequence.
ACCESSION
BM987874
VERSION
BM987874.1 GI:19707263
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 451)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://Image.llnl.gov>
Seq primer: M13 FORWARD
POLYA-yes.

FEATURES
source
1..451
Location/Qualifiers

```
FEATURES
source
location/Qualifiers
1..386
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1322779"
/clone_1lb="NCI_CGAP_K1d3
```

BASE COUNT	115 a	89 c	81 g	154 t	.
ORIGIN					
Query Match		13.08;	Score 49;	DB 9;	Length 439;

Best Local Similarity 100.0%; Pred. No. 2.9e-16;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CAGCATGTGTGCTAGTACGAGAGCTCTTGGCAAGATTCCTCC 346
|||||
Db 397 CAGCATGTGTGCTAGTACGAGAGCTCTTGGCAAGATTCCTCC 349

RESULT 4
LOCUS BG197212/c 806 bp mRNA linear EST 21-APR-2001
DEFINITION RST16451 Atherys RAGE Library Homo sapiens CDNA, mRNA sequence.
ACCESSION BG197212
VERSION BG197212.1 GI:13718899
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 806)
Harrington J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R., Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J., Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith, E., Veloso, N., Kikka, A., Hess, J., Colhoun, K., Lo, K., Offenbacher, J., Danzig, J. and Ducar, M.
Creation of genome-wide protein expression libraries using random activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
21227151

JOURNAL
MEDLINE Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9800
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 417.
Location/Qualifiers
1. 806

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-Wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 252 a 187 c 155 g 205 t 7 others
ORIGIN

Query Match 11.9%; Score 45; DB 12; Length 806;
Best Local Similarity 100.0%; Pred. No. 5.7e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 298 CAGCATGTGTGCTAGTACGAGAGCTCTTGGCAAGATTC 342
|||||
Db 387 CAGCATGTGTGCTAGTACGAGAGCTCTTGGCAAGATTC 343

RESULT 5
LOCUS AO035618 349 bp DNA linear GSS 11-JUL-1998
DEFINITION CIT-HSP-2319N3.TF CIT-HSP Homo sapiens genomic clone 2319N3, DNA sequence.
ACCESSION AO035618
VERSION AO035618.1 GI:3301715
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 349)

AUTHORS

TITLE

JOURNAL
COMMENT Unpublished (1998)
Other-GSSs: CIT-HSP-2319N3.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@ligr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.ligr.org/tcb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1. 349

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2319N3"
/clone_1lb="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelOBAC1; site_1: HindIII; site_2: HindIII"
BASE COUNT 85 a 69 c 84 g 111 t
ORIGIN

Query Match 9.0%; Score 34; DB 17; Length 349;
Best Local Similarity 100.0%; Pred. No. 7.9e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 86 CAGCCTGAATGCTGCTCGAGCTGAGAGATG 119
|||||
Db 31 CAGCCTGAATGCTGCTCGAGCTGAGAGATG 64

RESULT 6
LOCUS AO035738 237 bp DNA linear GSS 11-JUL-1998
DEFINITION CIT-HSP-2321C21.TF CIT-HSP Homo sapiens genomic clone 2321C21, DNA sequence.
ACCESSION AO035738
VERSION AO035738.1 GI:3301835
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 237)
Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Baas, S., Linher, K., Golden, K., Kounisley, S.D., Zhao, S., Field, C.E., Baas, S., Linher, K., Simon, M. and Venter, J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)
Unpublished (1998)
Other-GSSs: CIT-HSP-2321C21.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@ligr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.ligr.org/tcb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

JOURNAL
COMMENT Unpublished (1998)
Other-GSSs: CIT-HSP-2321C21.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@ligr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.ligr.org/tcb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

Unpublished (1998)
Other-GSSs: CIT-HSP-2321C21.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@ligr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.ligr.org/tcb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

```

FEATURES
  source
    Location/Qualifiers
      1..237
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_id="2321C21"
        /clone_lib="CIT-HSP"
        /sex="Male"
        /cell_type="Sperm"
        /note="Vector: pelobac11; site_1: HindIII; site_2: HindIII"
BASE COUNT      59 a      47 c      47 g      84 t
ORIGIN
Query Match      7.9%; Score 30; DB 17; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 348 GGAAGAGAGTACCATTGGAGAACCATGC 377
|||||
Db 224 GGAAGAGAGTACCATTGGAGAACCATGC 195

RESULT 7
LOCUS      BG210731      357 bp      mRNA      linear      EST 21-APR-2001
DEFINITION RST30277 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION  BG210731
VERSION     BG210731.1 GI:13732418
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 357)
  Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
  Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
  Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
  ,E., Veloso,N., Kikka,A., Hess,J., Cothren,K., Lo,K., Offenbacher
  ,J., Danzig,J. and Ducar,M.
  Creation of genome-wide protein expression libraries using random
  activation of gene expression
  Nat. Biotechnol. 19 (5), 440-445 (2001)
JOURNAL
MEDLINE     21227151
COMMENT     Contact: Scott J. Cain
  Athersys, Inc.
  3201 Carnegie Ave, Cleveland, OH 44115, USA
  Tel: 216 431 9900
  Fax: 216 361 9596
  Email: scaine@athersys.com
  High quality sequence stop: 357.
  Location/Qualifiers
    1..357
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="Athersys RAGE Library"
      /cell_line="HT1080"
      /note="See 'Creation of Genome-wide Protein Expression
  Libraries using Random Activation of Gene Expression',
  Nature Biotechnology, in press. Note that even though the
  cell type indicated is HT1080, since a random activation
  method was used, these sequence tags are not necessarily
  expressed in HT1080 under normal circumstances."
BASE COUNT      136 a      64 c      49 g      108 t
ORIGIN

Query Match      5.6%; Score 21; DB 12; Length 357;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 GGTGATGAGATGCATCAATCA 282
|||||
Db 40 GGTGATGAGATGCATCAATCA 20

FEATURES
  source
    Location/Qualifiers
      1..440
        /organism="Sus scrofa"
        /strain="crossbred"
        /db_xref="taxon:9823"
        /clone="MI-P-A2-afh-a-04-1-UM"
        /clone_lib="MI-P-A2"
        /lab_host="DH10B (Life Technologies)"
        /note="Vector: pT73D-Pac (Pharmacia) with a modified
  polylinker. Site 1: Not I; Site 2: EcoRI. The MI-P-A2
  library is derived from anterior pituitary at estrus day
  5. For a detailed description of the library from which
  this clone was derived, please visit our web site at
  http://piglet.genome.iastate.edu/.
  TAG_LIB=MI-P-A2
  TAG_TISSUE=4_5_anterior_pituitary
  TAG_SEQ=TCGCGT"
BASE COUNT      127 a      82 c      78 g      153 t
ORIGIN

Query Match      5.6%; Score 21; DB 14; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GAAAGAGGTACCATTGGAGA 370
|||||
Db 433 GAAAGAGGTACCATTGGAGA 413

RESULT 8
LOCUS      B0597872/c      440 bp      mRNA      linear      EST 24-JUN-2002
DEFINITION MI-P-A2-afh-a-04-1-UM.s1 MI-P-A2 Sus scrofa cDNA clone
  MI-P-A2-afh-a-04-1-UM 3', mRNA sequence.
ACCESSION  B0597872
VERSION     B0597872.1 GI:21544598
KEYWORDS
SOURCE      pig.
  Sus scrofa
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
REFERENCE   1 (bases 1 to 440)
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE     97044477
COMMENT     Contact: Tuggle CK
  Molecular Genetics Laboratory, Department of Animal Science
  Iowa State University
  201 Kildee Hall, Ames, IA 50011-3150, USA
  Tel: 5152944252
  Fax: 5152942401
  Email: cktuggle@iastate.edu
  Tissue Procurement: Dr. Chris Tuggle, Iowa State University
  cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Researchers may obtain clones from Research
  Genetics (www.resgen.com).
  The following repetitive elements were found in this cDNA
  sequence: 1-22, >AT-richLow_complexity
  Seq primer: M13 FORWARD
  POLY(A)-Yes.
  Location/Qualifiers
    1..440
      /organism="Sus scrofa"
      /strain="crossbred"
      /db_xref="taxon:9823"
      /clone="MI-P-A2-afh-a-04-1-UM"
      /clone_lib="MI-P-A2"
      /lab_host="DH10B (Life Technologies)"
      /note="Vector: pT73D-Pac (Pharmacia) with a modified
  polylinker. Site 1: Not I; Site 2: EcoRI. The MI-P-A2
  library is derived from anterior pituitary at estrus day
  5. For a detailed description of the library from which
  this clone was derived, please visit our web site at
  http://piglet.genome.iastate.edu/.
  TAG_LIB=MI-P-A2
  TAG_TISSUE=4_5_anterior_pituitary
  TAG_SEQ=TCGCGT"
BASE COUNT      127 a      82 c      78 g      153 t
ORIGIN

Query Match      5.6%; Score 21; DB 14; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GAAAGAGGTACCATTGGAGA 370
|||||
Db 433 GAAAGAGGTACCATTGGAGA 413

RESULT 9
LOCUS      BG223756      526 bp      mRNA      linear      EST 07-FEB-2001
DEFINITION 1M00018F10a Bovine Mixed Skeletal Muscle cDNA Library Bos taurus
  CDNA 5', mRNA sequence.
ACCESSION  BG223756
VERSION     BG223756.1 GI:12709277
KEYWORDS
  EST.

```

SOURCE
ORGANISM Bos taurus
 Cow.
REFERENCE Moore, S.S., Hansen, C., Li, C., Fu, A., Meng, Y., Li, G., Murdoch, G., Dixon, W. and Christopherson, B.
 1 (bases 1 to 526)
 CDNA's from bovine mixed skeletal muscle
 Unpublished (2001)
TITLE Beef Genomics Laboratory
JOURNAL Dept of AFNS, University of Alberta
COMMENT 410-Agril/For Centre, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
 Tel: 780 492 0169
 Fax: 780 492 4265
 Email: smoores@afns.ualberta.ca
 The sequence best matches gb:AF026293 (Homo sapiens chaperonin containing t-complex polypeptide 1, beta subunit (Cctb) mRNA, complete cds. 12/1998) in GenBank main database at E-value of 1e-113.
PCR PRIMERS
FORWARD: M13 Forward
BACKWARD: M13 Reverse
 Seq primer: T3 primer
 High quality sequence stop: 526
POLY-A-tails.

FEATURES
 source
 1..526
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="Bovine Mixed Skeletal Muscle cDNA Library"
 /sex="two males and one female mixed"
 /tissue_type="muscle, longissimuslorsal, biceps femoris, semimembranes"
 /cell_type="muscle"
 /dev_stage="Young adult"
 /lab_host="XLI-BlueMR-strain"
 /note="Organ: Skeletal muscle; Vector: Uni-22APXR; Site_1: Ecor I; Site_2: Xho I"
 Ecor I: 99 c 119 g 153 t

BASE COUNT 155 a 99 c 119 g 153 t
ORIGIN

Query Match 5.6%; Score 21; DB 12; Length 526;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GAAGGAGGTACCATGTGAGA 370
 ||||||||||||||||||
Db 111 GAAGGAGGTACCATGTGAGA 131

RESULT 10 541 bp mRNA linear EST 28-NOV-2001
LOCUS AV605455/c
DEFINITION AV605455 Bos taurus kidney fetus Bos taurus cDNA clone EIKI028H04
 3', mRNA sequence.
ACCESSION AV605455
VERSION AV605455.1 GI:9735828
KEYWORDS EST.
SOURCE EST.
ORGANISM Bos taurus
 Cow.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 541)
 Takasuga, A., Hirotsune, S., Itoh, R., Jitohzono, A., Suzuki, H., Aso, H. and Sugimoto, Y.
 Establishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 36,000 bovine ESTs
 Nucleic Acids Res. 29 (22), E108 (2001)

REFERENCE
AUTHORS
TITLE
JOURNAL

MEDLINE 21570554
COMMENT Contact: Yoshikazu Sugimoto
 Animal Genetics Division
 Shitakawa Institute of Animal Genetics
 Odakura, Nishigo, Nishi-shitakawa, Fukushima 961-8061, Japan
 Tel: 81-248-25-5641
 Fax: 81-248-25-5725
 Email: kazusugie@ocn.ne.jp
 Single pass sequencing.
 This clone was obtained from a polyA-deleted cDNA library.
FEATURES
 source
 1..541
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="EIKI028H04"
 /clone_lib="Bos taurus kidney fetus"
 /tissue_type="kidney"
 /dev_stage="fetus"
 /lab_host="DH10B"
 /note="Vector: pZ11; Site_1: SalI; Site_2: NotI; Poly A was deleted from a NotI site"
 139 a 136 c 110 g 153 t 3 others

BASE COUNT 139 a 136 c 110 g 153 t
ORIGIN

Query Match 5.6%; Score 21; DB 10; Length 541;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GAAGGAGGTACCATGTGAGA 370
 ||||||||||||||||||
Db 331 GAAGGAGGTACCATGTGAGA 311

RESULT 11 545 bp mRNA linear EST 28-NOV-2001
LOCUS AV605117/c
DEFINITION AV605117 Bos taurus kidney fetus Bos taurus cDNA clone EIKI028H04
 3', mRNA sequence.
ACCESSION AV605117
VERSION AV605117.1 GI:9735490
KEYWORDS EST.
SOURCE EST.
ORGANISM Bos taurus
 Cow.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 545)
 Takasuga, A., Hirotsune, S., Itoh, R., Jitohzono, A., Suzuki, H., Aso, H. and Sugimoto, Y.
 Establishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 36,000 bovine ESTs
 Nucleic Acids Res. 29 (22), E108 (2001)

JOURNAL 21570554
MEDLINE
COMMENT Contact: Yoshikazu Sugimoto
 Animal Genetics Division
 Shitakawa Institute of Animal Genetics
 Odakura, Nishigo, Nishi-shitakawa, Fukushima 961-8061, Japan
 Tel: 81-248-25-5641
 Fax: 81-248-25-5725
 Email: kazusugie@ocn.ne.jp
 Single pass sequencing.
 This clone was obtained from a polyA-deleted cDNA library.
FEATURES
 source
 1..545
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="EIKI028H04"
 /clone_lib="Bos taurus kidney fetus"
 /tissue_type="kidney"
 /dev_stage="fetus"
 /lab_host="DH10B"
 /note="Vector: pZ11; Site_1: SalI; Site_2: NotI; Poly A was deleted from a NotI site"

BASE COUNT 136 a 145 c 110 g 154 t
 ORIGIN was deleted from a NotI site"

Query Match 5.6%; Score 21; DB 10; Length 545;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 350 GAAGGAGGTACCATGGAGA 370
 DB 289 GAAGGAGGTACCATGGAGA 269

RESULT 12
 LOCUS BF713089 547 bp mRNA linear EST 02-JAN-2001
 DEFINITION MI-P-H3-adm-h-03-1-UM.s1 MI-P-H3 Sus scrofa cDNA clone
 ACCESSION BF713089
 VERSION BF713089
 KEYWORDS BF713089.1 GI:12012570
 SOURCE EST.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

REFERENCE 1 (bases 1 to 547)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 JOURNAL discovery
 MEDLINE Genome Res. 6 (9), 791-806 (1996)
 COMMENT Contact: Tuggle CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kluge Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152944201

FEATURES
 source
 Location/Qualifiers
 1..547
 /organism="Sus scrofa"
 /strain="crossbred"
 /db_xref="taxon:9623"
 /clone="MI-P-H3-adm-h-03-1-UM"
 /clone_1lb="MI-P-H3"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRTT3Pac (pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: EcoRI; The MI-P-H3
 library is derived from hypohalimus at estrus day 12.
 For a detailed description of the library from which this
 clone was derived, please visit our web site at
 http://pigest.genome.iastate.edu/
 TAG_lib=MI-P-H3
 TAG_TISSUE=hypohalimus at estrus day 12
 TAG_SEQ=GTAA
 TAG_SEQ=GTAA"

BASE COUNT 157 a 107 c 102 g 101 t
 ORIGIN

Query Match 5.6%; Score 21; DB 12; Length 547;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 350 GAAGGAGGTACCATGGAGA 370
 DB 450 GAAGGAGGTACCATGGAGA 430

RESULT 13
 LOCUS AM464207 569 bp mRNA linear EST 24-FEB-2000
 DEFINITION BP230015A10F9 Soares normalized bovine placenta Bos taurus cDNA
 ACCESSION AM464207
 VERSION AM464207
 KEYWORDS AM464207.1 GI:7034375
 SOURCE EST.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 569)
 Lewin,H.A., Soares,M.B., Rebelz,M., Pardinas,J., Liu,L. and Larson
 J.H.

REFERENCE 1
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 JOURNAL discovery
 MEDLINE Genome Res. 6 (9), 791-806 (1996)
 COMMENT Contact: Tuggle CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kluge Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152944201

FEATURES
 source
 Location/Qualifiers
 1..569
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone="BP230015A10F9"
 /clone_1lb="Soares normalized bovine placenta"
 /sex="female"
 /lab_host="DH10B"
 /note="Organ: Placenta; Vector: pRTT3Pac; Site_1: EcoRI;
 Site_2: NotI; The cDNA library was contributed by the
 Soares laboratory and it was constructed and normalized
 as described by Bonaldo, M.F., Lennon, G. and Soares,
 M.B. (1996), Genome Research 6(9): 791-806."
 BASE COUNT 160 a 112 c 144 g 150 t
 ORIGIN

Query Match 5.6%; Score 21; DB 10; Length 569;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 350 GAAGGAGGTACCATGGAGA 370
 DB 232 GAAGGAGGTACCATGGAGA 252

RESULT 14
 LOCUS A2364217 589 bp DNA linear GSS 02-OCT-2000

```

DEFINITION IM0110P18F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone uugc1m0110p18 f, DNA sequence.
ACCESSION A2364217
VERSION A2364217.1 GI:10477917
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 589)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0110 row: P column: 18
Seq primer: CGTTGTAAACGACGCCCACT
Class: plasmid ends
FEATURES
High quality sequence stop: 589.
location/Qualifiers
1..589
source
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0110P18"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-'"
/notes="Vector: pMD42uv, purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g11473211419b/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 130 a 141 c 110 g 208 t
ORIGIN
Query Match 5.6%; Score 21; DB 17; Length 589;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 219 CTTTGACTGTGAAAAGAGAC 239
|||||
Db 524 CTTTGACTGTGAAAAGAGAC 544

```

```

RESULT 15
B0600629/c 611 bp mRNA linear EST 24-JUN-2002
LOCUS B0600629

```

```

DEFINITION MI-P-E7-agz-d-10-1-UM.s1 MI-P-E7 Sus scrofa cDNA clone
MI-P-E7-agz-d-10-1-UM 3', mRNA sequence.
ACCESSION B0600629
VERSION B0600629.1 GI:21547355
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
1 (bases 1 to 611)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Tugue CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktugue@iastate.edu
Tissue Procurement: Dr. Chris Tugue, Iowa State University
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 1-22, >AT-rich/low-complexity
Seq primer: M13 FORWARD
POLYA=yes.
FEATURES
location/Qualifiers
1..611
source
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="MI-P-E7-agz-d-10-1-UM"
/clone_lib="MI-P-E7"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pSPORT1, site 1: Not I, site 2: EcoRI. The
MI-P-E7 library is derived from the following tissue(s):
prelongation day 12 conceptus. For a detailed description
of the library from which this clone was derived, please
visit our web site at http://pigst.genome.iastate.edu/.
TAG LIB-MI-P-E7
TAG TISSUE-prelongation_d_12_conceptus
TAG_SEQ-GTAGCA"
BASE COUNT 164 a 123 c 118 g 206 t
ORIGIN
Query Match 5.6%; Score 21; DB 14; Length 611;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 350 GAAGGAGGTACCATTTGGACA 370
|||||
Db 434 GAAGGAGGTACCATTTGGACA 414

```

```

Search completed: January 31, 2003, 05:14:00
Job time : 2258 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2003, 01:17:34 (Search time 267 seconds
(without alignments) 3168.222 Million cell updates/sec)

Title: US-09-803-719-222
Perfect score: 378
Sequence: 1 tacgcgtccttaagacgact.....tacacttgagagacatcg 378

Scoring table: OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

```
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	378	100.0	378	22 AAS37164	Novel human diagno
2	238	63.0	322	22 AAS37237	Novel human diagno
3	238	63.0	327	22 AAS37218	Novel human diagno
4	232	61.4	318	22 AAS37242	Novel human diagno
5	19	5.0	411	22 ABA20368	Human nervous syst
6	19	5.0	527	22 ABA61650	Human fetal liver
7	19	5.0	527	22 ABA29309	Human foetal liver
8	19	5.0	527	22 AAK09951	Human brain expres
9	19	5.0	527	22 AAK35845	Human bone marrow

10	19	5.0	527	22 AAI17155	Probe #7088 for ge
11	19	5.0	527	22 AAI15562	Probe #10248 used
12	19	5.0	527	22 ABA10078	Human genome-deriv
13	19	5.0	527	22 ABA17906	Human nervous syst
14	18	4.8	249	22 AAK58913	Human immune/haema
15	18	4.8	333	22 AAK74060	Human immune/haema
16	18	4.8	718	22 AAS62111	Porcine muscular s
17	18	4.8	721	24 AAS62084	Porcine muscular s
18	18	4.8	1768	23 ABL20677	Drosophila melanog
19	18	4.8	5090	23 ABL20676	Drosophila melanog
20	18	4.8	8638	23 ABL26654	Drosophila melanog
21	18	4.8	26555	22 AAK68372	Human immune/haema
22	18	4.8	26555	22 AAK68372	Human immune/haema
23	18	4.8	26555	22 AAK68372	Human immune/haema
24	17	4.5	281	21 AAK18996	Human immune/haema
25	17	4.5	305	22 AAK34310	Human secreted pro
26	17	4.5	597	23 AAV51017	Human haematologic
27	17	4.5	627	20 AAV8365	Human prostate exp
28	17	4.5	738	24 AAV8365	EST clone DD211.
29	17	4.5	740	22 AAV8365	M. capsulatus gene
30	17	4.5	783	21 AAV8365	Human neuroblastom
31	17	4.5	999	21 AAV8365	Arabidopsis thalia
32	17	4.5	1000	21 AAV8365	Human UGT1A7 relat
33	17	4.5	1001	21 AAV8365	Human UGT1A7 relat
34	17	4.5	1001	21 AAV8365	Human UGT1A7 relat
35	17	4.5	1001	21 AAV8365	Human UGT1A7 relat
36	17	4.5	1001	21 AAV8365	Human UGT1A7 relat
37	17	4.5	1041	22 AAK3200	Human foetal liver
38	17	4.5	1041	22 AAK3200	Human bone marrow
39	17	4.5	1041	22 AAK3200	Probe #17963 used
40	17	4.5	1041	22 AAK3200	Probe #9558 used t
41	17	4.5	1041	22 AAK3200	Human genome-deriv
42	17	4.5	1080	24 ABA91801	Staphylococcus epl
43	17	4.5	1180	23 ABA91801	Drosophila melanog
44	17	4.5	1293	21 AAK46337	Arabidopsis thalia
45	17	4.5	1294	21 AAK46337	Arabidopsis thalia

ALIGNMENTS

RESULT 1	
AA37164	AA37164 standard; cDNA; 378 BP.
AC	AA37164:
XX	
DT	17-DEC-2001 (first entry)
XX	
DE	Novel human diagnostic and therapeutic gene #222.
XX	
KW	Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO20016753-A2.
XX	
PD	13-SEP-2001.
XX	
PF	09-MAR-2001; 2001WO-US07787.
XX	
PR	09-MAR-2000; 2000US-0188609.
XX	
PA	(CHIR) CHIRON CORP.
XX	
PA	(HYSE) HYSEQ INC.
XX	
PI	Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI	Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI	Drimane R, Cirvenjakov R, Dickson M, Drimane S, Labat I;
PI	Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
DR	WPI: 2001-530177/58.
XX	

PT New polynucleotides and polypeptides, useful for diagnosis and
 PT treatment of breast, lung and colon cancer -

XX
 PS Claim 1: Page 652; 1193pp; English.

CC The invention relates to new polynucleotides and polypeptides, useful for
 CC diagnosis and treatment of breast, lung and colon cancer. The sequences
 CC can be used in detecting differentially expressed genes correlated with a
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample derived from a
 CC cell suspected of being cancerous. They can also be used to inhibit
 CC tumour growth by modulating expression of a gene product. AAS36943-
 CC AAS39338 represent novel human diagnostic and therapeutic coding
 CC sequences of the invention.

SO Sequence 378 BP; 100 A; 76 C; 101 G; 101 T; 0 other;

Query Match 100.0%; Score 378; DB 22; Length 378;
 Best Local Similarity 100.0%; Pred. No. 1,6e-187;

Matches 378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TACGGCTGCTTAAGACGCTTAAGGGGATGACGACGCGCTTTAGAGAACTATG 60
 Db 1 TACGGCTGCTTAAGACGCTTAAGGGGATGACGACGCGCTTTAGAGAACTATG 60
 QY 61 AAAACACCAAGCCGAGTCTCTCAACAAGTGAATGTGTCTGGAGCTGAAGGATG 120
 Db 61 AAAACACCAAGCCGAGTCTCTCAACAAGTGAATGTGTCTGGAGCTGAAGGATG 120
 QY 121 ACGGTTGTAAAGCCCGCTTTCTTCCGCTTTTAATCTAATGTTTGGAAATAAAAC 180
 Db 121 ACGGTTGTAAAGCCCGCTTTCTTCCGCTTTTAATCTAATGTTTGGAAATAAAAC 180
 QY 181 CTCCCTGCAAGTAGTACTTGTGTTTATGCTCAACATGCTTGAATGTTGAAAAGAGCC 240
 Db 181 CTCCCTGCAAGTAGTACTTGTGTTTATGCTCAACATGCTTGAATGTTGAAAAGAGCC 240
 QY 241 TTGGGACACATTGAAGGATGATGATGAGAGTCCCAATCCATGGAATCAAGTGGCGAG 300
 Db 241 TTGGGACACATTGAAGGATGATGATGAGAGTCCCAATCCATGGAATCAAGTGGCGAG 300
 QY 301 CTATGTTGATGATAGCAAGAGTCTTCTGGCAAGATTCCCTCCGGGAGAGAAAGTA 360
 Db 301 CTATGTTGATGATAGCAAGAGTCTTCTGGCAAGATTCCCTCCGGGAGAGAAAGTA 360
 QY 361 CCATTTGGAGAACCATGCG 378
 Db 361 CCATTTGGAGAACCATGCG 378

RESULT 2

AAS37237
 ID AAS37237 standard; cDNA; 322 BP.

AC AAS37237;

DT 17-DEC-2001 (first entry)

DE Novel human diagnostic and therapeutic gene #295.

KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.

OS Homo sapiens.

PN WO20016753-A2.

PD 13-SEP-2001.

PF 09-MAR-2001; 2001WO-US07787.

PR 09-MAR-2000; 2000US-0188609.

PA (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
 PI Dirmnac R, Crkvenjakov R, Dickson M, Dirmnac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
 DR WPI, 2001-530177/58.

PT New polynucleotides and polypeptides, useful for diagnosis and
 PT treatment of breast, lung and colon cancer -

PS Claim 1: Page 670; 1193pp; English.

CC The invention relates to new polynucleotides and polypeptides, useful for
 CC diagnosis and treatment of breast, lung and colon cancer. The sequences
 CC can be used in detecting differentially expressed genes correlated with a
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample derived from a
 CC cell suspected of being cancerous. They can also be used to inhibit
 CC tumour growth by modulating expression of a gene product. AAS36943-
 CC AAS39338 represent novel human diagnostic and therapeutic coding
 CC sequences of the invention.

SO Sequence 322 BP; 90 A; 64 C; 78 G; 89 T; 1 other;

Query Match 63.0%; Score 238; DB 22; Length 322;
 Best Local Similarity 100.0%; Pred. No. 2,6e-114;

Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 GGAAGAACCCCAAGCCGAGTCTCTCAACAGCTGAATGTTGTTGAGCTCAAGAT 118
 Db 67 GGAAGAACCCCAAGCCGAGTCTCTCAACAGCTGAATGTTGTTGAGCTCAAGAT 126
 QY 119 GCAGGTTGTTAAGCCCGTGTCTTCCGTTGTTTAACTAATGTTCTTGAATAAA 178
 Db 127 GCAGGTTGTTAAGCCCGTGTCTTCCGTTGTTTAACTAATGTTCTTGAATAAA 186
 QY 179 ACCTCCCTGCCAAGTAGTACTTGTGTTTATGCTCAACATGCTTGAATGTTGAAGA 238
 Db 187 ACCTCCCTGCCAAGTAGTACTTGTGTTTATGCTCAACATGCTTGAATGTTGAAGA 246
 QY 239 CTTTGGCACACATTGAAGGATGATGAGATGCAATCCATGGAATCAAGTGGC 296
 Db 247 CTTTGGCACACATTGAAGGATGATGAGATGCAATCCATGGAATCAAGTGGC 304

RESULT 3

AAS37218
 ID AAS37218 standard; cDNA; 327 BP.

AC AAS37218;

DT 17-DEC-2001 (first entry)

DE Novel human diagnostic and therapeutic gene #276.

KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.

OS Homo sapiens.

PN WO20016753-A2.

PD 13-SEP-2001.

PF 09-MAR-2001; 2001WO-US07787.

PR 09-MAR-2000; 2000US-0188609.

PA (CHIR) CHIRON CORP.

PI (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

PI Reinhard C, Randazzo F, Kennedy CC, Pot D, Kassam A, Lamson G;
 PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
 DR WPI: 2001-530177/58.

XX New polynucleotides and polypeptides, useful for diagnosis and
 PT treatment of breast, lung and colon cancer -
 XX

PS Claim 1: Page 665; 1193pp; English.

CC The invention relates to new polynucleotides and polypeptides, useful for
 CC diagnosis and treatment of breast, lung and colon cancer. The sequences
 CC can be used in detecting differentially expressed genes correlated with a
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample derived from a
 CC cell suspected of being cancerous. They can also be used to inhibit
 CC tumour growth by modulating expression of a gene product. AAS36943-
 CC AAS39338 represent novel human diagnostic and therapeutic coding
 CC sequences of the invention.

XX Sequence 327 BP; 92 A; 64 C; 81 G; 90 T; 0 other;

Query Match Best Local Similarity 63.0%; Score 238; DB 22; Length 327;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 GGAACACACCCAGCCGAGTCTCTCACAAGCTTGATGTGTTCTGAGCTGAAGAT 118
 DB 52 GGAACACACCCAGCCGAGTCTCTCACAAGCTTGATGTGTTCTGAGCTGAAGAT 111
 OY 119 GCAGCGTTGTTAAGCCCTGTTCTTTCCGTTGTTAATCTAATGTTCTTGGATTA 178
 DB 112 GCAGCGTTGTTAAGCCCTGTTCTTTCCGTTGTTAATCTAATGTTCTTGGATTA 171
 OY 179 ACCCTCCGCCAAGTAGTACTGTTTATGCTCAACATGCTTGACTGTGAAAAGAGA 238
 DB 172 ACCCTCCGCCAAGTAGTACTGTTTATGCTCAACATGCTTGACTGTGAAAAGAGA 231
 OY 239 CCTTGGCACAATGGAAGGATGATGATGAGATGCCAATCCATGGAATCAAGTGGC 296
 DB 232 CCTTGGCACAATGGAAGGATGATGATGAGATGCCAATCCATGGAATCAAGTGGC 289

RESULT 4
 ID AAS37242 standard; cDNA: 318 BP.

XX AAS37242;

DT 17-DEC-2001 (first entry)

DE Novel human diagnostic and therapeutic gene #300.

KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.

OS Homo sapiens.

PN WO200166753-A2.

PD 13-SEP-2001.

PF 09-MAR-2001; 2001WO-US07787.

PR 09-MAR-2000; 2000US-0188609.

PA (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
 PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;

XX WPI: 2001-530177/58.

XX New polynucleotides and polypeptides, useful for diagnosis and
 PT treatment of breast, lung and colon cancer -
 XX

PS Claim 1: Page 671; 1193pp; English.

CC The invention relates to new polynucleotides and polypeptides, useful for
 CC diagnosis and treatment of breast, lung and colon cancer. The sequences
 CC can be used in detecting differentially expressed genes correlated with a
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample derived from a
 CC cell suspected of being cancerous. They can also be used to inhibit
 CC tumour growth by modulating expression of a gene product. AAS36943-
 CC AAS39338 represent novel human diagnostic and therapeutic coding
 CC sequences of the invention.

XX Sequence 318 BP; 90 A; 63 C; 76 G; 89 T; 0 other;

Query Match Best Local Similarity 61.4%; Score 232; DB 22; Length 318;
 Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 GGAACACACCCAGCCGAGTCTCTCACAAGCTTGATGTGTTCTGAGCTGAAGAT 118
 DB 67 GGAACACACCCAGCCGAGTCTCTCACAAGCTTGATGTGTTCTGAGCTGAAGAT 126
 OY 119 GCAGCGTTGTTAAGCCCTGTTCTTTCCGTTGTTAATCTAATGTTCTTGGATTA 178
 DB 127 GCAGCGTTGTTAAGCCCTGTTCTTTCCGTTGTTAATCTAATGTTCTTGGATTA 186
 OY 179 ACCCTCCGCCAAGTAGTACTGTTTATGCTCAACATGCTTGACTGTGAAAAGAGA 238
 DB 187 ACCCTCCGCCAAGTAGTACTGTTTATGCTCAACATGCTTGACTGTGAAAAGAGA 246
 OY 239 CCTTGGCACAATGGAAGGATGATGATGAGATGCCAATCCATGGAATCA 290
 DB 247 CCTTGGCACAATGGAAGGATGATGATGAGATGCCAATCCATGGAATCA 298

RESULT 5
 ID ABA20368 standard; DNA: 411 BP.

XX ABA20368;

DT 23-JAN-2002 (first entry)

DE Human nervous system related polynucleotide SEQ ID NO 12699.

KW Human; nootropic; neuroprotective; cytostatic; dermatological; virocidic;
 KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnery;

KW antiparasitic; hepatotropic; antidiabetic; anticonvulsant; antifungal;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;

KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

PN Homo sapiens.

PD WO200159063-A2.

PF 16-AUG-2001.

PR 17-JAN-2001; 2001WO-US01334.

PR 04-FEB-2000; 2000US-0179065.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220963.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225367.
 PR 14-AUG-2000; 2000US-0225368.
 PR 14-AUG-2000; 2000US-0225370.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 18-AUG-2000; 2000US-0225759.
 PR 22-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226881.
 PR 22-AUG-2000; 2000US-0226886.
 PR 23-AUG-2000; 2000US-0227182.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 05-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0230433.
 PR 08-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0232081.
 PR 14-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 02-OCT-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 20-OCT-2000; 2000US-0242321.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246533.
 PR 08-NOV-2000; 2000US-0246534.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249289.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-541565/60.

Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 useful for preventing, diagnosing and/or treating nervous system
 cancers and metastases -

Disclosure: SEQ ID NO 12699; 1701pp + Sequence Listing; English.
 The invention relates to novel genes (ABA11004-ABA21534) and proteins
 (ABA14678-ABA18001) useful for preventing, treating or ameliorating
 medical conditions e.g. by protein or gene therapy. The genes are

CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SO Sequence 411 BP; 105 A; 79 C; 82 G; 145 T; 0 other;

Query Match 5.0%; Score 19; DB 22; Length 411;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 GTCTGGAGCTGAAGATG 119
 ||||||||||||||||
 DB 267 GTCTGGAGCTGAAGATG 285

RESULT 6

ABA61650
 ID ABA61650 standard; DNA: 527 BP.

AC ABA61650;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #9955.

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

PN WO200157277-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001MO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

DR WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human foetal liver -

PS Claim 1; SEQ ID NO 9955; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC foetal liver. The present sequence is a single exon nucleic acid

CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SO Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Query Match 5.0%; Score 19; DB 22; Length 527;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 TGCTTTGGATTAAC 180
 ||||||||||||||||
 DB 477 TGCTTTGGATTAAC 495

RESULT 7

ABA29309
 ID ABA29309 standard; DNA: 527 BP.

AC ABA29309;

DT 23-JAN-2002 (first entry)

DE Probe #7775 for gene expression analysis in human heart cell sample.

KW Human; gene expression; heart; microarray; vascular system; probe;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

XX congenital heart disease; ss.

OS Homo sapiens.

PN WO200157274-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001MO-US00666.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48899/53.

DR WPI; 2001-48899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human

XX hearts -

PS Claim 1; SEQ ID NO 7775; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for

CC measuring human gene expression in a sample derived from human heart. The

CC present sequence is one such probe. The probes may be used for

CC predicting, measuring and displaying gene expression in samples derived

CC from the human heart via microarrays. By measuring gene expression, the

CC probes are useful for predicting, diagnosing, grading, staging,

CC monitoring and prognosing diseases of the human heart and vascular system

CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and

CC congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SO Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Query Match 5.0%; Score 19; DB 22; Length 527;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 TGTCTTTGGAAATAAAC 180
 |||||
 DB 477 TGTCTTTGGAAATAAAC 495

RESULT 8

AAK09951
 ID AAK09951 standard; DNA; 527 BP.

AC AAK09951;
 XX

DT 05-NOV-2001 (first entry)
 XX

DE Human brain expressed single exon probe SEQ ID NO: 9942.
 XX

KM Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 XX epilepsy; cancer; ss.
 OS Homo sapiens.

OS Homo sapiens.
 XX

PN WO200157275-A2.
 XX

PD 09-AUG-2001.
 XX

PF 30-JAN-2001; 2001WO-US00667.
 XX

PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX

PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX

DR WPI; 2001-483446/52.
 XX

PT Single exon nucleic acid probes for analyzing gene expression in human
 XX brains -
 XX

PS Example 4; SEQ ID NO: 9942; 650bp + Sequence Listing; English.
 XX

CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system,
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention.
 XX

SO Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Query Match 5.0%; Score 19; DB 22; Length 527;
 Best Local Similarity 100.0%; Pred. No. 9;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 TGTCTTTGGAAATAAAC 180
 |||||
 DB 477 TGTCTTTGGAAATAAAC 495

RESULT 9

AAK35845
 ID AAK35845 standard; DNA; 527 BP.

AC AAK35845;
 XX

DT 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 10402.
 DE
 XX Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX
 OS Homo sapiens.

PN WO200157276-A2.
 XX

PD 09-AUG-2001.
 XX

PF 30-JAN-2001; 2001WO-US00668.
 XX

PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX

PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX

DR WPI; 2001-488900/53.
 XX

PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human bone marrow -
 XX

PS Example 4; SEQ ID NO: 10402; 658bp + Sequence Listing; English.
 XX

CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX

SO Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Query Match 5.0%; Score 19; DB 22; Length 527;
 Best Local Similarity 100.0%; Pred. No. 9;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 TGTCTTTGGAAATAAAC 180
 |||||
 DB 477 TGTCTTTGGAAATAAAC 495

RESULT 10

AAI17155
 ID AAI17155 standard; DNA; 527 BP.

AC AAI17155;
 XX

DT 12-OCT-2001 (first entry)
 XX

DE Probe #7088 for gene expression analysis in human cervical cell sample.
 XX

KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 XX

OS Homo sapiens.
 XX

PN WO200157278-A2.
 XX

PD 09-AUG-2001.
 XX

PF 30-JAN-2001; 2001WO-US00670.
 XX

PR 04-FEB-2000; 2000US-0180312.
 XX

PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI: 2001-488901/53.

PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human cervical epithelial cells -

PS Claim 25; SEQ ID No 7088; 487bp; English.

CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Query Match 5.0%; Score 19; DB 22; Length 527;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 TGTCTTGGAAATAAAC 180
 Db 477 TGTCTTGGAAATAAAC 495

RESULT 11

AA14562
 ID AA14562 standard; DNA: 527 BP.

XX AA14562;

DE 17-OCT-2001 (first entry)

XX Probe #10248 used to measure gene expression in human placenta sample.

KW Probe: microarray; human; placenta; antenatal diagnosis;
 KM genetic disorder; ss.

XX Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI: 2001-48897/53.

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -

PS Claim 25; SEQ ID No 10248; 654bp; English.

CC The present invention relates to single exon nucleic acid probes (SENPs).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.

SO Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Query Match 5.0%; Score 19; DB 22; Length 527;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 TGTCTTGGAAATAAAC 180
 Db 477 TGTCTTGGAAATAAAC 495

RESULT 12

ABS10078
 ID ABS10078 standard; DNA: 527 BP.

XX ABS10078;

DE 19-AUG-2002 (first entry)

XX Human genome-derived single exon probe from lung SEQ ID No 10069.

KW Human; ds: single exon probe; asthma; lung cancer; COPD; IID;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary hemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Kargener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.

XX Homo sapiens.

PN WO200186003-A2.

PD 15-NOV-2001.

PF 30-JAN-2001; 2001WO-US00665.

PR 04-FEB-2000; 2000US-180312P.

PR 26-MAY-2000; 2000US-207456P.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-234687P.

PR 27-SEP-2000; 2000US-236359P.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI: 2002-114183/15.

PT Spatially-addressable set of single exon nucleic acid probes, used to
 XX measure gene expression in human lung samples -

PS Claim 1; SEQ ID No 10069; 634bp; English.

CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived

CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridize at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC hemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Kartagener syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIP0 at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC XX

SQ Sequence 527 BP; 175 A; 99 C; 110 G; 143 T; 0 other;

Query Match 5.0%; Score 19; DB 24; Length 527;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 162 TGTCTTGGATTAAC 180
 |||||
 Db 477 TGTCTTGGATTAAC 495

RESULT 13
 ABA17906
 ID ABA17906 standard: DNA; 9929 BP.
 XX
 AC ABA17906:

DT 23-JAN-2002 (first entry)

DE Human nervous system related polynucleotide SEQ ID NO 10237.

XX
 KW Human; nootropic; neuroprotective; cyrostatic; dermatological; virocidic;
 KW immunosuppressive; antinflammatory; anti-HIV; antibacterial; vulnery;
 KW antiparinsonian; antischizoid; antianemic; antiairchic; cancer;
 KW antilethemic; hepatotropic; cerebroprotective; antinflammatory;
 KW antilethemic; antidiabetic; antilethemic; anticonvulsant; antifungal;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

OS Homo sapiens.
 XX
 PN WO200159063-A2.
 XX
 PD 16-AUG-2001.

XX
 PF 17-JAN-2001; 2001WO-0501334.
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0186874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214866.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 16-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 23-AUG-2000; 2000US-0227182.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 23-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.

```

PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251486.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0239678.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-541565/60.
XX

```

```

PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
XX Disclosure; SEQ ID NO 10237; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AB11004-AB121534) and proteins
CC (AB114678-AB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SO Sequence 9929 BP; 2748 A; 1678 C; 1979 G; 3524 T; 0 other;

Query Match 5.0%; Score 19; DB 22; Length 9929;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 TTAAGCCCGCTGCTTTTC 146
Db 7124 TTAAGCCCGCTGCTTTTC 7142
|||||
|||||

RESULT 14
ID AAK58913/c
ID AAK58913 standard; cDNA; 249 BP.
XX
AC AAK58913;
XX
DT 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:3973.
DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:3973.
KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW Cytostatic; gene therapy; vaccine; metastasis; ss.
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 28-JUN-2000; 2000US-0209467.
XX 30-JUN-2000; 2000US-0214886.
XX 07-JUL-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 11-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 14-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.

```

PR	26-JUL-2000	2000US-0220966
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225413
PR	14-AUG-2000	2000US-0225516
PR	14-AUG-2000	2000US-0225614
PR	14-AUG-2000	2000US-0225667
PR	14-AUG-2000	2000US-0225668
PR	14-AUG-2000	2000US-0225679
PR	14-AUG-2000	2000US-0225750
PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	14-AUG-2000	2000US-0225765
PR	14-AUG-2000	2000US-0225779
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226688
PR	22-AUG-2000	2000US-0227182
PR	22-AUG-2000	2000US-0227189
PR	30-ANG-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229293
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229513
PR	05-SEP-2000	2000US-0229519
PR	05-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230348
PR	06-SEP-2000	2000US-0231422
PR	08-SEP-2000	2000US-0231343
PR	08-SEP-2000	2000US-0231344
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0231480
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231868
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0232398
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	14-SEP-2000	2000US-0233066
PR	14-SEP-2000	2000US-0233225
PR	21-SEP-2000	2000US-0234273
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236127
PR	29-SEP-2000	2000US-0236167
PR	29-SEP-2000	2000US-0236568
PR	29-SEP-2000	2000US-0236569
PR	29-SEP-2000	2000US-0236570
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239335
PR	13-OCT-2000	2000US-0239337
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241877
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	01-NOV-2000	2000US-0241826
PR	08-NOV-2000	2000US-0246417
PR	08-NOV-2000	2000US-0246418

[illegible]

CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 249 BP; 68 A; 64 C; 40 G; 74 T; 3 other;

Query Match 4.8%; Score 18; DB 22; Length 249;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 257 GGGATGCTGATGAGATG 274
|||||
DB 78 GCGATGCTGATGAGATG 61

RESULT 15
AAK74060/c
ID AAK74060 standard; DNA; 333 BP.
XX
AC AAK74060;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:28872.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN W0200157182-A2.
XX
PD 09-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216860.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225265.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2003, 02:17:29 ; Search time 2723 Seconds

(without alignments)
4039,978 Million cell updates/sec

Title: US-09-803-719-222

Sequence: 1 tacgcgtcctaagacgact.....taccatggaagaccatgacg 378

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_srs:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_or:*

21: em_ov:*

22: em_pat:*

23: em_ph:*

24: em_pl:*

25: em_ro:*

26: em_srs:*

27: em_un:*

28: em_vl:*

29: em_hum:*

30: em_hum:*

31: em_hum:*

32: em_hum:*

33: em_hum:*

34: em_hum:*

35: em_hum:*

36: em_hum:*

37: em_hum:*

38: em_hum:*

39: em_hum:*

40: em_hum:*

41: em_hum:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	378	100.0	378	6	AX245292 Sequence
2	238	63.0	322	6	AX245365 Sequence
3	238	63.0	327	6	AX245346 Sequence
4	232	61.4	318	6	AX245370 Sequence
5	119	31.5	59231	9	AL135937 Human DNA
6	20	5.3	65118	2	AC084187 Homo sapi
7	20	5.3	65118	9	AL451145 Human DNA
8	20	5.3	100403	2	AP003087 Homo sapi
9	20	5.3	122370	2	AP004044 Oryza sat
10	20	5.3	129591	2	AP005251 Oryza sat
11	20	5.3	131682	9	AL672277 Human DNA
12	20	5.3	132557	9	AC004520 Homo sapi
13	20	5.3	156469	2	AC046149 Mus muscu
14	20	5.3	159108	10	AC009295 Mus muscu
15	20	5.3	162394	2	AC119775 Rattus no
16	20	5.3	170371	9	AP001132 Homo sapi
17	20	5.3	171638	10	AC124708 Mus muscu
18	20	5.3	187983	2	AC026210 Homo sapi
19	20	5.3	187987	2	AC009639 Homo sapi
20	20	5.3	191141	2	AC069374 Homo sapi
21	20	5.3	202251	2	AC104472 Homo sapi
22	20	5.3	268444	2	AC110152 Homo sapi
23	20	5.0	6469	9	AB007865 Homo sapi
24	19	5.0	6854	1	BS039230 Homo sapi
25	19	5.0	7260	9	AF339163 Cynocypa
26	19	5.0	56703	2	AC100286 Mus muscu
27	19	5.0	60648	5	AL672044 zebrafish
28	19	5.0	93939	1	AE006471 Salmonell
29	19	5.0	106607	9	AL355259 Human DNA
30	19	5.0	113810	2	AC068292 Homo sapi
31	19	5.0	130468	2	CNS08CAE Homo sapi
32	19	5.0	141818	2	AL133389 Homo sapi
33	19	5.0	142522	9	AC090625 Homo sapi
34	19	5.0	155132	2	AC095530 Rattus no
35	19	5.0	158475	9	AC097463 Homo sapi
36	19	5.0	161731	2	AC128360 Rattus no
37	19	5.0	164074	2	AC103987 Homo sapi
38	19	5.0	168114	9	AL354828 Human DNA
39	19	5.0	169886	2	AC016578 Homo sapi
40	19	5.0	170334	9	AC016579 Homo sapi
41	19	5.0	175395	2	AC119119 Rattus no
42	19	5.0	177147	9	AC007539 Homo sapi
43	19	5.0	181300	9	HSBA512M4 Human DNA
44	19	5.0	181413	2	AP001352 Homo sapi
45	19	5.0	188132	9	AC079304 Homo sapi

ALIGNMENTS

RESULT 1	AX245292	378 bp	DNA	linear	PAT 28-SEP-2001
LOCUS	AX245292				
DEFINITION	Sequence 222 from Patent WO0166753.				
ACCESSION	AX245292				
VERSION	AX245292.1				
KEYWORDS	GI:15859966				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
	Williams, L.T., Escobedo, J., Innis, M.A., Garcia, P.D.,				
	Sudduth-Klinger, J., Reinhard, C., Randazzo, F., Kennedy, G.C., Pot, D.,				
	Kassam, A., Lamson, G., Drmanac, R., Crkvenjakov, R., Dickson, M.,				

TITLE Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
Stache-Crain,B.
JOURNAL Human genes and gene expression products
Patent: WO 0166753-A 222 13-SEP-2001;
Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES
source 1..378
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 100 a 76 c 101 g 101 t

Query Match 100.0%; Score 378; DB 6; Length 378;
Best Local Similarity 100.0%; Pred. No. 4,9e-225;
Matches 378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGCGCTGTTAAGACACTTAAGGGGATGACGCGCGCTTGAAGAACATATG 60
|||||
Db 1 TAGCGCTGTTAAGACACTTAAGGGGATGACGCGCGCTTGAAGAACATATG 60
QY 61 AAACACCCAGCCGAGTCTCTCACAGCTTGATGTGTTGAGCTGAAGATGC 120
|||||
Db 61 AAACACCCAGCCGAGTCTCTCACAGCTTGATGTGTTGAGCTGAAGATGC 120
QY 121 AGCGTTGTAAGCCCTGTTCTTTCGTTTATCTAATGTTCTTTGGATTAAC 180
|||||
Db 121 AGCGTTGTAAGCCCTGTTCTTTCGTTTATCTAATGTTCTTTGGATTAAC 180
QY 181 CTCCTGCGCAAGTACTGTTGTTTATGCTCAACATGCTTTGACTGTTGAAAAGACC 240
|||||
Db 181 CTCCTGCGCAAGTACTGTTGTTTATGCTCAACATGCTTTGACTGTTGAAAAGACC 240
QY 241 TTTGGCACACATTGAAGGATGTGATGAGATGCCAATCCATGGAATCAGTGGCCAG 300
|||||
Db 241 TTTGGCACACATTGAAGGATGTGATGAGATGCCAATCCATGGAATCAGTGGCCAG 300
QY 301 CTATGTTGGTAGTATAGCAGAGTCTTCTTGCAAGATTCCTCCGCGGAAGAGATGA 360
|||||
Db 301 CTATGTTGGTAGTATAGCAGAGTCTTCTTGCAAGATTCCTCCGCGGAAGAGATGA 360
QY 361 CCATTGAGAACCATGCG 378
|||||
Db 361 CCATTGAGAACCATGCG 378

RESULT 2
AX245365 322 bp DNA linear PAT 28-SEP-2001
LOCUS AX245365 Sequence 295 from Patent WO0166753.
DEFINITION AX245365
ACCESSION AX245365
VERSION AX245365.1 GI:15860039
KEYWORDS
SOURCE human;
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 322)
Mammalia: Eutheria: Chordata: Craniata: Vertebrata: Euteleostomi:
Aukaryota: Eutheria: Primates: Catarrhini: Homini: Homo.
AUTHORS Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
Kassam,A., Lamson,G., Drmanac,R., Crivenjakov,R., Dickson,M.,
Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
Stache-Crain,B.

TITLE Human genes and gene expression products
JOURNAL Patent: WO 0166753-A 295 13-SEP-2001;
Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES
source 1..322
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 90 a 64 c 78 g 89 t 1 others
ORIGIN
Query Match 63.0%; Score 238; DB 6; Length 322;

Best Local Similarity 100.0%; Pred. No. 3.2e-137;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 GGAACACCCAGCCGAGTCTCTCACAGCTTGAATGTGTGAGCTGAAGAT 118
|||||
Db 67 GGAACACCCAGCCGAGTCTCTCACAGCTTGAATGTGTGAGCTGAAGAT 126
QY 119 GCACGGTGTAAAGCCCTGTTCTTTCGTTTATCTAATGTTCTTTGGAATAAA 178
|||||
Db 127 GCACGGTGTAAAGCCCTGTTCTTTCGTTTATCTAATGTTCTTTGGAATAAA 186
QY 179 ACCCTCCGCAAGTACTGTTGTTTATGCTCAACATGCTTGACTGTTGAAAAGAGA 238
|||||
Db 187 ACCCTCCGCAAGTACTGTTGTTTATGCTCAACATGCTTGACTGTTGAAAAGAGA 246
QY 239 CCTTTGGCACACATTGAAGGATGTGATGAGATGCCAATCCATGGAATCAGGTGC 296
|||||
Db 247 CCTTTGGCACACATTGAAGGATGTGATGAGATGCCAATCCATGGAATCAGGTGC 304

RESULT 3
AX245346 327 bp DNA linear PAT 28-SEP-2001
LOCUS AX245346 Sequence 276 from Patent WO0166753.
DEFINITION AX245346
ACCESSION AX245346
VERSION AX245346.1 GI:15860020
KEYWORDS
SOURCE human;
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 327)
Mammalia: Eutheria: Primates: Catarrhini: Homini: Homo.
AUTHORS Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
Kassam,A., Lamson,G., Drmanac,R., Crivenjakov,R., Dickson,M.,
Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
Stache-Crain,B.

TITLE Human genes and gene expression products
JOURNAL Patent: WO 0166753-A 276 13-SEP-2001;
Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES
source 1..327
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 92 a 64 c 81 g 90 t

Query Match 63.0%; Score 238; DB 6; Length 327;
Best Local Similarity 100.0%; Pred. No. 3.2e-137;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 GGAACACCCAGCCGAGTCTCTCACAGCTTGAATGTGTGAGCTGAAGAT 118
|||||
Db 52 GGAACACCCAGCCGAGTCTCTCACAGCTTGAATGTGTGAGCTGAAGAT 111
QY 119 GCACGGTGTAAAGCCCTGTTCTTTCGTTTATCTAATGTTCTTTGGAATAAA 178
|||||
Db 112 GCACGGTGTAAAGCCCTGTTCTTTCGTTTATCTAATGTTCTTTGGAATAAA 171
QY 179 ACCCTCCGCAAGTACTGTTGTTTATGCTCAACATGCTTGACTGTTGAAAAGAGA 238
|||||
Db 172 ACCCTCCGCAAGTACTGTTTATGCTCAACATGCTTTGACTGTTGAAAAGAGA 231
QY 239 CCTTTGGCACACATTGAAGGATGTGATGAGATGCCAATCCATGGAATCAGGTGC 296
|||||
Db 232 CCTTTGGCACACATTGAAGGATGTGATGAGATGCCAATCCATGGAATCAGGTGC 289

RESULT 4
AX245370 318 bp DNA linear PAT 28-SEP-2001
LOCUS AX245370 Sequence 300 from Patent WO0166753.
DEFINITION AX245370
ACCESSION AX245370

VERSION AX245370.1 GI:15860044
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 318)
 AUTHORS Williams, L.T., Escobedo, J., Innis, M.A., Garcia, P.D., Sudduth-Klinger, J., Reinhard, C., Randazzo, F., Kennedy, G.C., Pot, D., Kaasam, A., Lamson, G., Drmanac, R., Kravetsky, R., Dickson, M., Drmanac, S., Labat, I., Leshkowitz, D., Kita, D., Garcia, Y. and Steche-Crain, B.
 TITLE Human genes and gene expression products
 JOURNAL Patent: WO 0166753-A 300 13-SEP-2001;
 Chiron Corporation (US); Hyseq Inc. (US)
 FEATURES
 source 1..318
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 90 a 63 c 76 g 89 t
 ORIGIN
 Query Match 61.4%; Score 232; DB 6; Length 318;
 Best Local Similarity 100.0%; Pred. No. 1.8e-133;
 Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 59 GGAAACACCAAGCCGAGTCTCTCAAGCTGATGTGTCTGAGCTGAAGAT 118
 Db 67 GGAAACACCAAGCCGAGTCTCTCAAGCTGATGTGTCTGAGCTGAAGAT 126
 Oy 119 GCACGGTGTGAAGCCCTGTTTCCCTGTTTAACTAATGTTCTTGGAAATAAA 178
 Db 127 GCACGGTGTGAAGCCCTGTTTCCCTGTTTAACTAATGTTCTTGGAAATAAA 186
 Oy 179 ACTCCCTGCACATGTACTGTTTATGTCACACATCTTGTGAGCTGGAAGAAGA 238
 Db 187 ACTCCCTGCACATGTACTGTTTATGTCACACATCTTGTGAGCTGGAAGAAGA 246
 Oy 239 CCTTTGGCACATTTGAAGGATGTGTGAGATGCCAATTCATGGAATCA 290
 Db 247 CCTTTGGCACATTTGAAGGATGTGTGAGATGCCAATTCATGGAATCA 298
 RESULT 5
 AL135937 59231 bp DNA linear PRI 15-MAR-2001
 LOCUS Human DNA sequence from clone RPI-278022 on chromosome 20. Contains
 DEFINITION STS and GSSs, complete sequence.
 ACCESSION AL135937
 VERSION AL135937.22 GI:7248202
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 59231)
 AUTHORS Phillimore, B.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 COMMENT On Mar 15, 2000 this sequence version replaced GI:7242336. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Ch20>
 IMPORTANT: This sequence is not the entire insert of clone RPI-278022. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RPI-278022 is at 59132 in this sequence. The true right end of clone RPI-278022 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RPI-278022 is from the library RPI-1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECT: PCYFAC2.
 FEATURES
 source Location/Qualifiers
 1..59231
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="RPI-278022"
 /clone_1lb="RPI-1"
 58..213
 /note="MER5A repeat: matches 9. .188 of consensus"
 233..409
 /note="MER3 repeat: matches 6. .191 of consensus"
 572..795
 /note="MER repeat: matches 3. .241 of consensus"
 2927..2992
 /note="A1u repeat: matches 242. .307 of consensus"
 3019..3635
 /note="LIME2 repeat: matches 5541. .6154 of consensus"
 3880..4411
 /note="LIM4 repeat: matches 4015. .4580 of consensus"
 4412..4600
 /note="MER58A repeat: matches 32. .224 of consensus"
 complement(5199..5709)
 /note="match: GSS: Em:AQ613263"
 7115..7585
 /note="Charlieda repeat: matches 25. .507 of consensus"
 7631..7695
 /note="MER repeat: matches 13. .81 of consensus"
 8546..9223
 /note="LIMC4 repeat: matches 7117. .7739 of consensus"
 9986..10209
 /note="A1uSG repeat: matches 1. .225 of consensus"
 10225..10250
 /note="13 copies 2 mer aa 96% conserved"
 complement(10684..11081)
 /note="match: GSS: Em:AQ62878"
 11093..11168
 /note="38 copies 2 mer aa 72% conserved"
 11575..12292
 /note="LIMAG repeat: matches 5545. .6278 of consensus"
 12860..13177
 /note="match: GSS: Em:AQ035618"
 13327..13423
 /note="L2 repeat: matches 2657. .2750 of consensus"
 13719..13813
 /note="L2 repeat: matches 2195. .2293 of consensus"
 13814..14087
 /note="LIMAS repeat: matches 6025. .6298 of consensus"
 14088..14219
 /note="L2 repeat: matches 2078. .2195 of consensus"
 14965..15067
 /note="MER repeat: matches 169. .262 of consensus"
 complement(15295..15803)
 /note="match: GSS: Em:AQ754141"

```

misc_feature      15849..16248
                   /note="match: GSS: Em:AQ0694331"
repeat_region     15961..16498
                   /note="LIME3A repeat: matches 5427..5964 of consensus"
repeat_region     16499..16808
                   /note="AlusX repeat: matches 1..310 of consensus"
repeat_region     16809..17004
                   /note="LIME3A repeat: matches 5964..6159 of consensus"
repeat_region     17356..17544
                   /note="MIR repeat: matches 26..230 of consensus"
repeat_region     18267..18357
                   /note="MIR repeat: matches 50..139 of consensus"
repeat_region     18358..18690
                   /note="MER58B repeat: matches 2..341 of consensus"
repeat_region     18691..18717
                   /note="MIR repeat: matches 24..50 of consensus"
repeat_region     18885..18952
                   /note="34 copies 2 mer ga 79% conserved"
misc_feature      complement(19364..19600)
                   /note="match: GSS: Em:AQ035738"
repeat_region     22132..22242
                   /note="LIR33 repeat: matches 410..521 of consensus"
repeat_region     22243..22551
                   /note="AlusX repeat: matches 1..310 of consensus"
repeat_region     22552..22951
                   /note="LIR33 repeat: matches 17..410 of consensus"
repeat_region     23356..23643
                   /note="AlusG repeat: matches 1..288 of consensus"
repeat_region     24124..24510
                   /note="LIR16C repeat: matches 1..387 of consensus"
repeat_region     24747..24810
                   /note="32 copies 2 mer ta 68% conserved"
repeat_region     24967..25054
                   /note="LIM1 repeat: matches 5678..5763 of consensus"
repeat_region     25134..25443
                   /note="AlusG repeat: matches 1..313 of consensus"
repeat_region     25767..26069
                   /note="AlusG repeat: matches 1..304 of consensus"
repeat_region     26293..26579
                   /note="LIMC/D repeat: matches 5290..5579 of consensus"
repeat_region     27383..27481
                   /note="MIR repeat: matches 20..142 of consensus"
repeat_region     28338..28469
                   /note="LIMC4 repeat: matches 7841..7973 of consensus"
repeat_region     28472..28583
                   /note="56 copies 2 mer ta 73% conserved"
repeat_region     28584..28970
                   /note="LIMC4 repeat: matches 7430..7849 of consensus"
repeat_region     30290..30876
                   /note="L2 repeat: matches 2124..2658 of consensus"
repeat_region     30877..31223
                   /note="MER1B repeat: matches 1..337 of consensus"
repeat_region     31224..31421
                   /note="L2 repeat: matches 1945..2124 of consensus"
repeat_region     31422..31739
                   /note="AluY repeat: matches 1..309 of consensus"
repeat_region     31750..32046
                   /note="AlusG repeat: matches 3..298 of consensus"
repeat_region     32047..32188
                   /note="L2 repeat: matches 1809..1937 of consensus"
repeat_region     32189..32474
                   /note="AluJo repeat: matches 1..293 of consensus"
repeat_region     32475..32713
                   /note="L2 repeat: matches 1420..1809 of consensus"
repeat_region     34859..34938
                   /note="MIR repeat: matches 191..261 of consensus"
repeat_region     35160..35351
                   /note="MER5A repeat: matches 7..189 of consensus"
repeat_region     36121..36273
                   /note="MIR repeat: matches 66..225 of consensus"
repeat_region     37904..37941
                   /note="19 copies 2 mer aa 81% conserved"
repeat_region     38788..38840

```

```

repeat_region     /note="MIR repeat: matches 37..84 of consensus"
repeat_region     38841..39254
                   /note="MIR repeat: matches 1..426 of consensus"
repeat_region     39255..39304
                   /note="MIR repeat: matches 84..150 of consensus"
repeat_region     39310..39339
                   /note="LIM3d repeat: matches 251..280 of consensus"
repeat_region     39314..39717
                   /note="LIM3e repeat: matches -879..779 of consensus"
repeat_region     39724..39990
                   /note="LIM9 repeat: matches 3735..4011 of consensus"
repeat_region     39992..40126
                   /note="AluJo repeat: matches 2..134 of consensus"
repeat_region     40153..40379
                   /note="LIM9 repeat: matches 3999..4226 of consensus"
repeat_region     40380..40675
                   /note="AluJo repeat: matches 1..306 of consensus"
repeat_region     40676..42726
                   /note="LIM9 repeat: matches 4226..6308 of consensus"
repeat_region     43333..43524
                   /note="LIR16C repeat: matches 165..386 of consensus"
misc_feature      complement(43484..44004)
                   /note="match: GSS: Em:AQ0800109"
repeat_region     44765..44818
                   /note="MIR repeat: matches 200..250 of consensus"
misc_feature      complement(45503..46337)

```

```

Query Match      31.58; Score 119; DB 9; Length 59231;
Best local Similarity 100.0%; Pred. No. 1,6e-62;
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 148 GTGCTTAATCATATGTCCTTGGAAATAAACCTCCCTGCAGTACTAGTGGTTTA 207
      |||||||
Db 18738 GTGCTTAATCATATGTCCTTGGAAATAAACCTCCCTGCAGTACTAGTGGTTTA 18797
      |||||||

Qy 208 TGCTCAACATGCTTGCAGTGGAAAGAGACCTTGGCACATTTGAAGGATGCTGA 266
      |||||||
Db 18798 TGCTCAACATGCTTGCAGTGGAAAGAGACCTTGGCACATTTGAAGGATGCTGA 18856
      |||||||

```

```

RESULT 6
LOCUS      AC084187
DEFINITION Homo sapiens clone RP11-609K10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC084187
VERSION    AC084187.1 GI:10801415
KEYWORDS   HTG; HTGS_PHA5E0.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 65118)
            Birren,B., Linton,L., Nusbaum,C. and Lander,E.
            Homo sapiens, Clone RP11-609K10
            Unpublished
            2 (bases 1 to 65118)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Batra,N., Bastien,V., Beda,F., Boguslavsky,L.,
            Boukhalter,B., Brown,A., Burkett,G., Campolino,A., Castle,A.,
            Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
            Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Glade,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heatford,A., Horton,L.,
            Iller,I., Johnson,R., Jones,C., Kann,L., Karakas,A., Lacroque,K.,
            Lamaizares,R., Landers,T., Lehotzky,J., Levine,R., Liu,C., Liu,G.,
            Macdonald,P., Margulis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheters,R., Meldrum,J., Menus,L., Mihova,T., Mlenga,Y.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rottman,R., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,

```

TITLE
JOURNAL
COMMENT

Firrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
Zimmer, A. and Zody, M.
Submitted (14-Oct-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L11460

Center clone name: 609_K_10

NOTE: This record contains 77 individual

* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 747 846: contig of 746 bp in length
* 747 846: gap of 100 bp
* 847 1605: contig of 759 bp in length
* 1606 1705: gap of 100 bp
* 1706 2465: contig of 760 bp in length
* 2466 2565: gap of 100 bp
* 2566 3313: contig of 748 bp in length
* 3314 3413: gap of 100 bp
* 3414 4145: contig of 732 bp in length
* 4146 4245: gap of 100 bp
* 4246 4953: contig of 708 bp in length
* 4954 5053: gap of 100 bp
* 5054 5818: contig of 765 bp in length
* 5819 5918: gap of 100 bp
* 5919 6664: contig of 746 bp in length
* 6665 6764: gap of 100 bp
* 6765 7496: contig of 732 bp in length
* 7497 7596: gap of 100 bp
* 7597 8344: contig of 748 bp in length
* 8345 8444: gap of 100 bp
* 8445 9198: contig of 754 bp in length
* 9199 9298: gap of 100 bp
* 9299 10047: contig of 749 bp in length
* 10048 10147: gap of 100 bp
* 10148 10902: contig of 755 bp in length
* 10903 11002: gap of 100 bp
* 11003 11785: contig of 783 bp in length
* 11786 11885: gap of 100 bp
* 11886 12636: contig of 751 bp in length
* 12637 12736: gap of 100 bp
* 12737 13503: contig of 767 bp in length
* 13504 13603: gap of 100 bp
* 13604 14357: contig of 754 bp in length
* 14358 14457: gap of 100 bp
* 14458 15208: contig of 751 bp in length
* 15209 15308: gap of 100 bp
* 15309 16026: contig of 718 bp in length
* 16027 16126: gap of 100 bp
* 16127 16867: contig of 741 bp in length
* 16868 16967: gap of 100 bp
* 16968 17697: contig of 730 bp in length
* 17698 17797: gap of 100 bp
* 17798 18538: contig of 741 bp in length
* 18539 18638: gap of 100 bp

18639 19403: contig of 765 bp in length
* 19404 19503: gap of 100 bp
* 19504 20248: contig of 745 bp in length
* 20249 20348: gap of 100 bp
* 20349 21101: contig of 753 bp in length
* 21102 21201: gap of 100 bp
* 21202 21932: contig of 731 bp in length
* 21933 22032: gap of 100 bp
* 22033 22790: contig of 758 bp in length
* 22791 22890: gap of 100 bp
* 22891 23641: contig of 751 bp in length
* 23642 23741: gap of 100 bp
* 23742 24460: contig of 739 bp in length
* 24461 24580: gap of 100 bp
* 24581 25321: contig of 741 bp in length
* 25322 25421: gap of 100 bp
* 25422 26169: contig of 748 bp in length
* 26170 26269: gap of 100 bp
* 26270 27016: contig of 747 bp in length
* 27017 27116: gap of 100 bp
* 27117 27881: contig of 765 bp in length
* 27882 27981: gap of 100 bp
* 27982 28747: contig of 766 bp in length
* 28748 28847: gap of 100 bp
* 28848 29607: contig of 760 bp in length
* 29608 29707: gap of 100 bp
* 29708 30457: contig of 750 bp in length
* 30458 30557: gap of 100 bp
* 30558 31317: contig of 760 bp in length
* 31318 31417: gap of 100 bp
* 31418 32167: contig of 750 bp in length
* 32168 32267: gap of 100 bp
* 32268 33014: contig of 747 bp in length
* 33015 33114: gap of 100 bp
* 33115 33850: contig of 736 bp in length
* 33851 33950: gap of 100 bp
* 33951 34698: contig of 748 bp in length
* 34699 34798: gap of 100 bp
* 34799 35541: contig of 743 bp in length
* 35542 35641: gap of 100 bp
* 35642 36392: contig of 751 bp in length
* 36393 36492: gap of 100 bp
* 36493 37242: contig of 750 bp in length
* 37243 37342: gap of 100 bp
* 37343 38091: contig of 749 bp in length
* 38092 38191: gap of 100 bp
* 38192 38923: contig of 732 bp in length
* 38924 39023: gap of 100 bp
* 39024 39777: contig of 754 bp in length
* 39778 39877: gap of 100 bp
* 39878 40651: contig of 774 bp in length
* 40652 40751: gap of 100 bp
* 40752 41468: contig of 717 bp in length
* 41469 41568: gap of 100 bp
* 41569 42287: contig of 719 bp in length
* 42288 42387: gap of 100 bp
* 42388 43133: contig of 746 bp in length
* 43134 43233: gap of 100 bp
* 43234 43970: contig of 737 bp in length
* 43971 44070: gap of 100 bp
* 44071 44814: contig of 744 bp in length
* 44815 44914: gap of 100 bp
* 44915 45672: contig of 758 bp in length
* 45673 45772: gap of 100 bp
* 45773 46509: contig of 737 bp in length
* 46510 46609: gap of 100 bp
* 46610 47344: contig of 735 bp in length
* 47345 47444: gap of 100 bp
* 47445 48184: contig of 740 bp in length
* 48185 48284: gap of 100 bp
* 48285 49047: contig of 763 bp in length
* 49048 49147: gap of 100 bp
* 49148 49884: contig of 737 bp in length

Query Match 5.3% Score 20; DB 2; Length 65118;
 Best Local Similarly 100.0%; Pred. No. 2;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 172 AATNAAACCTCCTGCCAA 191
 |||||||||||||||||||
 Db 53064 AATNAAACCTCCTGCCAA 53083

RESULT 7
 AL451145/c 65811 bp DNA linear PRI 08-JUN-2001
 LOCUS Human DNA sequence from clone RP11-164A17 on chromosome 6, complete
 DEFINITION

ACCESSION AL451145
 VERSION AL451145.15 GI:14348489
 KEYWORDS htc.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 65811)
 Garner, P.

REFERENCE 2
 Direct Submision
 Submitted (08-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 Request: clonerequest@sanger.ac.uk
 On Jun 11, 2001 this sequence version replaced gi:14268205.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw: S;
 SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
 Group. Further information can be found at

FEATURES
 source
 http://www.sanger.ac.uk/HGP/Chr6
 RP11-164A17 is from the library RPc1-11.1 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/dacpac/home.htm>
 VECTOR: pBACE3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-164A17 it may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP3-365019 is at 65712 in this sequence.
 The true right end of clone RP11-54515 is at 100 in this sequence.

1..65811
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="RP11-164A17"
 /clone_1ib="RPC1-11.1"
 complement(1..114)
 /note="match: GSS: Em:A0744388"
 complement(1..100)
 /note="match: GSS: Em:A0351948"
 220..352
 /note="MER5A repeat: matches 55..183 of consensus"
 2175..2344
 /note="MER2 repeat: matches 2..165 of consensus"
 3016..3163
 /note="L2 repeat: matches 139..295 of consensus"
 3464..3765
 /note="L2 repeat: matches 1242..1583 of consensus"
 3805..4497
 /note="L2 repeat: matches 1713..2479 of consensus"
 4519..4552
 /note="L2 repeat: matches 1713..2479 of consensus"
 4555..5939
 /note="L2 repeat: matches 1713..2479 of consensus"
 5944..6122
 /note="L2 repeat: matches 1713..2479 of consensus"
 6146..6787
 /note="match: GSS: Em:A0473725"
 complement(6757..7094)
 /note="match: GSS: Em:B33281"
 7878..8157
 /note="L2 repeat: matches 1..285 of consensus"
 10517..10648
 /note="MER46A repeat: matches 1..134 of consensus"
 10814..11006
 /note="L2 repeat: matches 1..134 of consensus"
 11139..11355
 /note="L2 repeat: matches 2..223 of consensus"
 11397..11750
 /note="L2 repeat: matches 2..223 of consensus"
 11801..11976
 /note="L2 repeat: matches 2580..2953 of consensus"
 12076..12353
 /note="88 copies 2 mer ta 61% conserved"
 12076..12353
 /note="L2 repeat: matches 1778..2060 of consensus"
 complement(12258..12657)
 /note="match: GSS: Em:A0076115"
 12429..12969
 /note="L2 repeat: matches 1201..1743 of consensus"
 12981..13114
 /note="MER46A repeat: matches 1..142 of consensus"
 13512..13792
 /note="L2 repeat: matches 1..304 of consensus"
 13908..14156
 /note="L2 repeat: matches 1..304 of consensus"
 14242..14678
 /note="match: GSS: Em:A0825235"
 complement(15008..15441)
 /note="match: GSS: Em:A0391327"
 complement(15964..20391)
 /note="match: GSS: Em:A0244751"
 20006..20049
 /note="22 copies 2 mer at 77% conserved"

AP004044/c 122370 bp DNA linear HTG 21-MAR-2002
 LOCUS Oryza sativa (japonica cultivar-group) chromosome 8 clone
 DEFINITION OJ1125.C01, *** SEQUENCING IN PROGRESS ***; in ordered pieces.
 ACCESSION AP004044.1 GI:15208412
 VERSION HTG; HTGS_PHASE2.
 KEYWORDS Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
 SOURCE clone:OJ1125.C01.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 REFERENCE 1 Sasaki,T., Matsumoto,T. and Yamamoto,K.
 AUTHORS Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, BAC
 TITLE clone:OJ1125.C01
 JOURNAL Published only in Database (2001)
 REFERENCE 2 (bases 1 to 122370)
 AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
 JOURNAL Direct Submission
 TITLE Submitted (15-AUG-2001) Takuji Sasaki, National Institute of
 JOURNAL Agricultural Resources, Rice Genome Research Program; Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasakientas.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 COMMENT The nucleotide sequence of this BAC clone was generated by
 combining Monsanto and RGP-Japan sequencing data.
 NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source
 1. 122370
 /organism="Oryza sativa (japonica cultivar-group)"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="8"
 /clone="OJ1125.C01"
 BASE COUNT 36453 a 25424 c 26087 g 34354 t 52 others
 ORIGIN
 Query Match 5.3%; Score 20; DB 2; Length 122370;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 TACGGCTGCTTAAGACGACT 20
 ||||||||||||||||||
 Db 121735 TACGGCTGCTTAAGACGACT 121716
 RESULT 10
 AP005251 129591 bp DNA linear HTG 24-MAY-2002
 LOCUS Oryza sativa (japonica cultivar-group) chromosome 8 clone
 DEFINITION OSJNB0011H15, *** SEQUENCING IN PROGRESS ***; in ordered pieces.
 ACCESSION AP005251.1 GI:21165567
 VERSION HTG; HTGS_PHASE2.
 KEYWORDS Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
 SOURCE clone:OSJNB0011H15.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 REFERENCE 1 Sasaki,T., Matsumoto,T. and Katayose,Y.

TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, BAC
 clone:OSJNB0011H15
 JOURNAL Published only in Database (2002)
 REFERENCE 2 (bases 1 to 129591)
 AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
 JOURNAL Direct Submission
 TITLE Submitted (23-MAY-2002) Takuji Sasaki, National Institute of
 JOURNAL Agricultural Sciences, Rice Genome Research Program; Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasakientas.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 COMMENT NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source
 1. 129591
 /organism="Oryza sativa (japonica cultivar-group)"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="8"
 /clone="OSJNB0011H15"
 BASE COUNT 36092 a 27555 c 28367 g 37453 t 124 others
 ORIGIN
 Query Match 5.3%; Score 20; DB 2; Length 129591;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 TACGGCTGCTTAAGACGACT 20
 ||||||||||||||||||
 Db 16209 TACGGCTGCTTAAGACGACT 16190
 RESULT 11
 AL672277 131682 bp DNA linear PRI 20-AUG-2002
 LOCUS Human DNA sequence from clone RP11-309M23 on chromosome X, complete
 DEFINITION sequence.
 ACCESSION AL672277.20 GI:22449770
 VERSION HTG.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 131682)
 AUTHORS Clark,S.
 JOURNAL Direct Submission
 TITLE Submitted (20-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Aug 22, 2002 this sequence version replaced g1:21911559.
 COMMENT ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate


```

repeat_region      /rpt_family="Alu"
4508. .5230
/rpt_family="ERV1"
repeat_region      /rpt_family="Alu"
5323. .5630
/rpt_family="Alu"
repeat_region      /rpt_family="L2"
6328. .6422
/rpt_family="L2"
repeat_region      /rpt_family="MIR"
6937. .7144
/rpt_family="MIR"
repeat_region      /rpt_family="L2"
7409. .7490
/rpt_family="L2"
repeat_region      /rpt_family="Alu"
7708. .7953
/rpt_family="Alu"
repeat_region      /rpt_family="Alu"
7954. .8040
/rpt_family="Alu"
repeat_region      /rpt_family="Alu"
8272. .8569
/rpt_family="Alu"
repeat_region      /rpt_family="MIR"
8622. .8696
/rpt_family="MIR"
repeat_region      /rpt_family="ERV1"
8708. .9164
/rpt_family="ERV1"
repeat_region      /rpt_family="MIR"
9180. .9274
/rpt_family="MIR"
repeat_region      /rpt_family="MIR"
11518. .11717
/rpt_family="MIR"
repeat_region      /rpt_family="MIR"
11718. .11817
/rpt_family="MIR"
repeat_region      /rpt_family="Alu"
13943. .14241
/rpt_family="Alu"
repeat_region      /rpt_family="Alu"
14463. .14687
/rpt_family="Alu"
repeat_region      /rpt_family="Alu"
15533. .15784
/rpt_family="Alu"
repeat_region      /rpt_family="L1"
15785. .15908
/rpt_family="L1"
repeat_region      /rpt_family="Alu"
15909. .15986
/rpt_family="Alu"
repeat_region      /rpt_family="L1"
16143. .16300
/rpt_family="L1"
repeat_region      /rpt_family="L1"
16480. .16528
/rpt_family="L1"
repeat_region      /rpt_family="MIR"
16706. .16987
/rpt_family="MIR"
repeat_region      /rpt_family="Alu"
17881. .18181
/rpt_family="Alu"
repeat_region      /rpt_family="Alu"
18493. .18796
/rpt_family="Alu"
repeat_region      /rpt_family="Alu"
19290. .19470
/rpt_family="Alu"
repeat_region      /rpt_family="MIR"
20605. .20824
/rpt_family="MIR"
repeat_region      /rpt_family="ERV1"
20825. .21126
/rpt_family="ERV1"
repeat_region      /rpt_family="Alu"
21127. .21529
/rpt_family="Alu"
repeat_region      /rpt_family="ERV1"
22556. .22787
/rpt_family="ERV1"
repeat_region      /rpt_family="MIR"
22795. .22839
/rpt_family="MIR"
repeat_region      /rpt_family="MIR"
22860. .23070
/rpt_family="MIR"
repeat_region      /rpt_family="MIR"
23851. .24034
/rpt_family="MIR"
repeat_region      /rpt_family="MIR"
24038. .24353
/rpt_family="MIR"
repeat_region      /rpt_family="Alu"
25234. .25546
/rpt_family="Alu"
repeat_region      /rpt_family="Alu"
26003. .26451
/rpt_family="Alu"
repeat_region      /rpt_family="ERV1"
26626. .26914
/rpt_family="ERV1"
repeat_region      /rpt_family="Alu"
27468. .27532
/rpt_family="Alu"

```

```

repeat_region      28341. .28410
/rpt_family="L2"
repeat_region      28546. .28691
/rpt_family="MIR"
repeat_region      29423. .29500
/rpt_family="MIR"
repeat_region      29635. .29987
/rpt_family="MIR"
repeat_region      30261. .30406
/rpt_family="MIR"
repeat_region      30407. .30601
/rpt_family="Alu"
repeat_region      30602. .30765
/rpt_family="Alu"
repeat_region      31508. .31716
/rpt_family="Alu"
repeat_region      33321. .33438
/rpt_family="L1"
repeat_region      33600. .33809
/rpt_family="MIR"
repeat_region      /rpt_family="L1"

```

```

Query Match      5.38; Score 20; DB 9; Length 132557;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      84 CACAAGCTGCAATGTGTT 103
Db      13715 CACAAGCTGCAATGTGTT 13696

```

```

RESULT 13
AC046149/c      156469 bp      DNA      linear      HTG 16-OCT-2001
LOCUS      MUS musculus chromosome 10 clone RP23-310B11, WORKING DRAFT
DEFINITION      AC046149
ACCESSION      AC046149
VERSION      AC046149.5 GI:16118092
KEYWORDS      HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE      Mus musculus.
ORGANISM      Mus musculus.

```

```

REFERENCE
AUTHORS      Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Bahay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gili,R.,
Gorell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Louissege,H., Lozano,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstein,G.,
Worley,K. and Gibbs,R.
TITLE      Unpublished
JOURNAL      Direct Submission
REFERENCE      2 (bases 1 to 156469)
AUTHORS      Worley,K.C.
TITLE      Direct Submission
JOURNAL      Submitted (13-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Oct 14, 2001 this sequence version replaced gi:11094642.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

```



```

repeat_region 2048. .2155 /rpt_family="GC_rich"
repeat_region 2228. .2290 /rpt_family="GC_rich"
repeat_region complement(7061. .7239) /rpt_family="AluDo"
repeat_region complement(7448. .7557) /rpt_family="FLM_A"
repeat_region 7581. .7618 /rpt_family="(CAAAA)n"
repeat_region 11048. .11067 /rpt_family="(GGGA)n"
repeat_region 11346. .11393 /rpt_family="(CA)n"
repeat_region 12783. .12816 /rpt_family="(TTTA)n"
repeat_region complement(12904. .12963) /rpt_family="MIR"
repeat_region 14619. .14709 /rpt_family="FLM_C"
repeat_region 15382. .15475 /rpt_family="(TCTC)n"
repeat_region complement(15532. .15625) /rpt_family="AluDb"
repeat_region 16802. .16860 /rpt_family="AluSp"
repeat_region 17130. .17310 /rpt_family="(CATTA)n"
repeat_region 20035. .20128 /rpt_family="FLM_C"
repeat_region 20387. .20435 /rpt_family="(CA)n"
repeat_region 20436. .20528 /rpt_family="(GA)n"
repeat_region 20837. .20867 /rpt_family="(TGG)n"
repeat_region 21617. .21641 /rpt_family="AT_rich"
repeat_region 22773. .22821 /rpt_family="AT_rich"
repeat_region 22878. .23043 /rpt_family="(TTTAA)n"
repeat_region 23211. .23273 /rpt_family="FRAM"
repeat_region 23448. .23475 /rpt_family="CT_rich"
repeat_region 24116. .24201 /rpt_family="(TGG)n"
repeat_region 24359. .24450 /rpt_family="AluDo"
repeat_region 25402. .25429 /rpt_family="LMB2"
repeat_region 25429. .25494 /rpt_family="(TGG)n"
repeat_region 26461. .26508 /rpt_family="(GA)n"
repeat_region 27167. .27197 /rpt_family="(CA)n"
repeat_region 27210. .27269 /rpt_family="AT_rich"
repeat_region 27633. .27673 /rpt_family="MIR"
repeat_region 28793. .28877 /rpt_family="(TGG)n"
repeat_region 28921. .28958 /rpt_family="AluS"
repeat_region complement(29020. .29094) /rpt_family="MIR"
repeat_region 31216. .31256 /rpt_family="CT_rich"
repeat_region complement(31814. .32710) /rpt_family="LMA5A"
repeat_region 33264. .33314 /rpt_family="LMA5A"

```

```

/rpt_family="HERV15I|HERV15I"
35155. .35191 /rpt_family="AT_rich"
repeat_region 35254. .35300 /rpt_family="(GGGGA)n"
repeat_region 35725. .35764 /rpt_family="(TGG)n"
repeat_region 38626. .38803 /rpt_family="(CA)n"
repeat_region 39467. .39493 /rpt_family="AT_rich"
repeat_region 40008. .40048 /rpt_family="(TTTTC)n"
repeat_region complement(40052. .40170) /rpt_family="5S"
repeat_region 40298. .40417 /rpt_family="(GGA)n"
repeat_region complement(42703. .42930) /rpt_family="LMA6"
repeat_region 43475. .43511 /rpt_family="(TCTG)n"
repeat_region 43513. .43596 /rpt_family="(TGG)n"
repeat_region 43596. .43672 /rpt_family="(GA)n"
repeat_region complement(44004. .44081) /rpt_family="LMA5A"
repeat_region 48910. .48930 /rpt_family="(TGG)n"
repeat_region complement(49480. .49658) /rpt_family="AluDo"
repeat_region 51864. .51890 /rpt_family="AT_rich"
repeat_region 52072. .52212 /rpt_family="FRAM"
repeat_region 52299. .52326 /rpt_family="(TCTCC)n"
repeat_region 52414. .52521 /rpt_family="AluS"
repeat_region 52664. .52950 /rpt_family="AluDo"
repeat_region 52951. .52980 /rpt_family="(CCAA)n"
repeat_region 52981. .52995 /rpt_family="AluDo"
repeat_region complement(54449. .54619) /rpt_family="AluDo"
repeat_region 58720. .58753 /rpt_family="(TGG)n"

Query Match 5.3%: Score 20: DB 10: Length 159108;
Best Local Similarity 100.08; Pred. No. 2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 344 TCCCGGAGAGAGGTACCA 363
DB 105501 TCCCGGAGAGAGGTACCA 105482

RESULT 15
AC119775/c 162394 bp DNA linear HTG 18-JUL-2002
LOCUS Rattus norvegicus clone CH230-464J18, *** SEQUENCING IN PROGRESS
DEFINITION *** 61 unordered pieces.
ACCESSION AC119775
VERSION AC119775.4 GI:21747184
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 162394)

```

AUTHORS

MUZYU,D.M., Adams,C., Adio-Obiola,B., Ali-osman,F.R., Allen,C., Alsproucks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbatia,U., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,U., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douhwalte,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Fritztz,P., Gabisi,A., Geo,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homel,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lileu,C., Liu,J., Liu,W., Louleaged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nokenkwo,S., Oguh,M., Okwono,G., Ogunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L., Qulies,M., Ren,Y., Rivers,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Severy,G., Scherer,S., Scott,G., Shen,H., Shoohteril,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Vatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansy,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalona,D., Vinson,R., Wang,O., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

TITLE

Unpublished
2 (bases 1 to 162394)

REFERENCE

Worley,K.C.

Direct Submission

Submitted (02-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 162394)

Worley,K.C.

Direct Submission

Submitted (18-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jul 14, 2002 this sequence version replaced gi:20467838.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GYQM

Center clone name: CH230-464U18

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye 100% of reads

Assembly program: Phrap: version 0.990329

Consensus quality: 104933 bases at least Q40

Consensus quality: 112584 bases at least Q20

Consensus quality: 117408 bases at least Q20

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently consists of 61 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1004: contig of 1004 bp in length
1005 1104: gap of unknown length
1105 2121: contig of 1017 bp in length
2122 2221: gap of unknown length
2222 3241: contig of 1020 bp in length
3242 3341: gap of unknown length
3342 4689: contig of 1348 bp in length
4690 4789: gap of unknown length
4790 6424: contig of 1633 bp in length
6425 8023: gap of unknown length
8024 8123: contig of 1499 bp in length
8124 9619: gap of unknown length
9619 9718: contig of 1495 bp in length
9718 10759: contig of 1041 bp in length
10760 10859: gap of unknown length
10860 12638: contig of 1779 bp in length
12639 12738: gap of unknown length
12739 14083: contig of 1345 bp in length
14084 14183: gap of unknown length
14184 15891: contig of 1708 bp in length
15892 15991: gap of unknown length
15992 17623: contig of 1632 bp in length
17624 17723: gap of unknown length
17724 18962: contig of 1239 bp in length
18963 19062: gap of unknown length
19063 20849: contig of 1787 bp in length
20850 20949: gap of unknown length
20950 22169: contig of 1220 bp in length
22170 22269: gap of unknown length
22270 22483: contig of 1214 bp in length
22484 23583: gap of unknown length
23584 24875: contig of 1292 bp in length
24876 24975: gap of unknown length
24976 26670: contig of 1695 bp in length
26671 26770: gap of unknown length
26771 27914: contig of 1144 bp in length
27915 28014: gap of unknown length
28015 29124: contig of 1110 bp in length
29125 29224: gap of unknown length
29226 31386: contig of 2162 bp in length
31387 31486: gap of unknown length
31487 32500: contig of 1014 bp in length
32501 32600: gap of unknown length
32601 33435: contig of 1745 bp in length
33436 34445: gap of unknown length
34446 36597: contig of 2152 bp in length
36598 36697: gap of unknown length
36698 38768: contig of 2071 bp in length
38769 38868: gap of unknown length
38869 40546: contig of 1678 bp in length
40547 40646: gap of unknown length
40647 41770: contig of 1124 bp in length
41771 41870: gap of unknown length
41871 43735: contig of 1865 bp in length
43736 43835: gap of unknown length
43836 45206: contig of 1771 bp in length
45207 45306: gap of unknown length
45307 47814: contig of 2508 bp in length
47815 49406: contig of 1492 bp in length
49407 49506: gap of unknown length
49507 51910: contig of 2404 bp in length
51911 52010: gap of unknown length
52011 54105: contig of 2095 bp in length

```
* 54106 54205: gap of unknown length
* 54206 56954: contig of 2749 bp in length
* 56955 57054: gap of unknown length
* 57055 59393: contig of 2339 bp in length
* 59394 59493: gap of unknown length
* 59494 61754: contig of 2261 bp in length
* 61755 61854: gap of unknown length
* 61855 64055: contig of 2201 bp in length
* 64056 64155: gap of unknown length
* 64156 67129: contig of 2974 bp in length
* 67130 67229: gap of unknown length
* 67230 69622: contig of 2393 bp in length
* 69623 69723: gap of unknown length
* 69723 72435: contig of 2713 bp in length
* 72436 72535: gap of unknown length
* 72536 75742: contig of 3207 bp in length
* 75743 75842: gap of unknown length
* 75843 77891: contig of 2049 bp in length
* 77892 77991: gap of unknown length
* 77992 80486: contig of 2495 bp in length
* 80487 80586: gap of unknown length
* 80587 83986: contig of 3400 bp in length
* 83987 84086: gap of unknown length
* 84087 86359: contig of 2273 bp in length
* 86360 86459: gap of unknown length
* 86460 91249: contig of 4790 bp in length
* 91250 91349: gap of unknown length
* 91350 94806: contig of 3457 bp in length
* 94807 94906: gap of unknown length
* 94907 98994: contig of 4088 bp in length
* 98995 99094: gap of unknown length
* 99095 105047: contig of 5953 bp in length
* 105048 105147: gap of unknown length
* 105148 108668: contig of 3521 bp in length
* 108669 108768: gap of unknown length
* 108769 111829: contig of 3061 bp in length
* 111830 111929: gap of unknown length
* 111930 116397: contig of 4468 bp in length
* 116398 116497: gap of unknown length
* 116498 122101: contig of 5604 bp in length
* 122102 122201: gap of unknown length
* 122202 126352: contig of 4151 bp in length
```

```
Query Match 5.3%: Score 20; DB 2; Length 162394;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 210 CTCACATGCTTGACTGT 229
Db 79021 CTCACATGCTTGACTGT 79002
```

Search completed: January 31, 2003, 04:46:58
Job time : 3388 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2003, 01:07:09 : Search time 206 Seconds
(without alignments)
2082.401 Million cell updates/sec

Title: US-09-803-719-222

Perfect score: 378
Sequence: 1 tacggctgctaagaagcact.....tacattggaacacatgcg 378

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2475311 seqs, 567426730 residues

Total number of hits satisfying chosen parameters: 4950622

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_NA_New :
1: /cgnt2_6/ptodata/1/pna/PCT_NEW_COMB.seq : *
2: /cgnt2_6/ptodata/1/pna/US06_NEW_COMB.seq : *
3: /cgnt2_6/ptodata/1/pna/US07_NEW_COMB.seq : *
4: /cgnt2_6/ptodata/1/pna/US08_NEW_COMB.seq : *
5: /cgnt2_6/ptodata/1/pna/US09_NEW_COMB.seq : *
6: /cgnt2_6/ptodata/1/pna/US10_NEW_COMB.seq : *
7: /cgnt2_6/ptodata/1/pna/US60_NEW_COMB.seq : *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35.2	9.3	1671	PCT-US02-10366-23	Sequence 23, Appl
2	35.2	9.3	1677	PCT-US02-10366-21	Sequence 21, Appl
3	35.2	9.3	2067	PCT-US02-10366-27	Sequence 27, Appl
4	35.2	9.3	2130	PCT-US02-10366-25	Sequence 25, Appl
5	33.6	8.9	492	US-10-276-774-299	Sequence 299, Appl
6	33.6	8.9	1629	PCT-US02-38526-589	Sequence 589, Appl
7	33.6	8.9	1640	PCT-US02-38526-587	Sequence 587, Appl
8	33.4	8.8	158	US-09-513-999C-35882	Sequence 33882, A
9	32.6	8.6	45450	PCT-US02-32700-7	Sequence 7, Appl1
10	32.6	8.6	45450	US-10-271-413-5935	Sequence 5935, Ap
11	32.4	8.5	334	US-09-531-116-5935	Sequence 5935, Ap
12	32.2	8.5	264	US-09-950-084-952	Sequence 952, Ap
13	32	8.5	410	US-09-513-999C-30304	Sequence 30304, A
14	31.6	8.4	2966	US-10-294-443-103	Sequence 103, App
15	31.6	8.4	3011	US-10-294-443-112	Sequence 112, App
16	31.6	8.4	3701	US-10-294-443-110	Sequence 110, Appl
17	30.8	8.1	1083	US-09-724-676-6004	Sequence 6004, Ap
18	30.8	8.1	1083	US-09-724-676-6004	Sequence 6004, Ap
19	30.8	8.1	1180	US-09-724-676-6035	Sequence 6035, Ap
20	30.8	8.1	1180	US-09-724-676-6035	Sequence 6035, Ap
21	30.8	8.1	1208	US-10-276-774-1101	Sequence 1101, Ap
22	30.8	8.1	1392	US-09-724-676-5990	Sequence 5990, Ap
23	30.8	8.1	1392	US-09-724-676-5990	Sequence 5990, Ap
24	30.8	8.1	1489	US-09-724-676-6021	Sequence 6021, Ap
25	30.8	8.1	1489	US-09-724-676-6021	Sequence 6021, Ap
26	30.8	8.1	1770	US-09-724-676-5980	Sequence 5980, Ap

C 27	30.8	8.1	1770	5	US-09-724-676A-5980	Sequence 5980, Ap
C 28	30.8	8.1	1867	5	US-09-724-676-6011	Sequence 6011, Ap
C 29	30.8	8.1	1867	5	US-09-724-676A-6011	Sequence 6011, Ap
C 30	30.4	8.0	2347	6	US-10-240-425-389	Sequence 389, App
C 31	30.4	8.0	2771	5	US-09-724-676-37523	Sequence 37523, A
C 32	30.4	8.0	2771	5	US-09-724-676A-37523	Sequence 37523, A
C 33	30.4	8.0	2866	5	US-09-724-676-37524	Sequence 37524, A
C 34	30.4	8.0	2866	5	US-09-724-676A-37524	Sequence 37524, A
C 35	30.2	8.0	36303	6	US-10-152-724-24	Sequence 24, Appl
C 36	30.2	8.0	36303	6	US-10-152-724A-24	Sequence 24, Appl
C 37	30.2	8.0	302603	1	PCT-US02-32700-8	Sequence 8, Appl1
C 38	30.2	8.0	302603	1	US-10-271-416-8	Sequence 1, Appl1
C 39	30	7.9	14759	6	US-10-346-723-1	Sequence 131, App
C 40	30	7.9	26159	6	US-10-144-779-131	Sequence 230, App
C 41	30	7.9	75829	6	US-10-144-779-230	Sequence 389, App
C 42	29.8	7.9	472	5	US-09-531-113-31549	Sequence 31549, A
C 43	29.8	7.9	1457	6	US-10-289-757-41	Sequence 41, Appl
C 44	29.8	7.9	859705	5	US-09-948-124-60	Sequence 60, Appl
C 45	29.6	7.8	648	5	US-09-513-999C-3839	Sequence 3839, Ap

ALIGNMENTS

RESULT 1
PCT-US02-10366-23
Sequence 23, Application PC/TUS0210366
GENERAL INFORMATION:
APPLICANT: Padigar, Muralidhara
APPLICANT: Shenoy, Suresh
APPLICANT: Kekuda, Ramesh
APPLICANT: Rastelli, Luca
APPLICANT: Mezes, Peter
APPLICANT: Smithson, Glenda
APPLICANT: Guo, Xiaojia
APPLICANT: Gerlach, Valerie
APPLICANT: Casman, Stacie
APPLICANT: Boldog, Ferenc
APPLICANT: Li, Li
APPLICANT: Zernusen, Bryan
APPLICANT: Tchernev, Velizar
APPLICANT: Gangoli, Eshe
APPLICANT: Verneet, Corine
APPLICANT: Spytek, Kimberly
APPLICANT: Malyankar, Uriei
APPLICANT: Patlurajan, Meera
APPLICANT: Miller, Charles
APPLICANT: Taupier, Raymond J. Jr.
APPLICANT: Hayes, Melvyn
APPLICANT: Ju, Jingfang
APPLICANT: Peyman, John
APPLICANT: Carteron, Elina
APPLICANT: MacDougall, John
APPLICANT: Edinger, Shlomit
APPLICANT: Stone, David
APPLICANT: Mazur, Ann
TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES,
TITLE OF INVENTION: ACIDS ENCODING THE ANTIGENS, AND METHODS OF USE
FILE REFERENCE: 21402-322A-061
CURRENT APPLICATION NUMBER: PCT/US02/10366
CURRENT FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: 10/114153
PRIOR FILING DATE: 2002-04-02
PRIOR APPLICATION NUMBER: 60/281086
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/281906
PRIOR FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 60/282020
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: 60/282930
PRIOR FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: 60/283512

```
;; PRIOR FILING DATE: 2001-04-12
;; PRIOR APPLICATION NUMBER: 60/283444
;; PRIOR FILING DATE: 2001-04-12
;; PRIOR APPLICATION NUMBER: 60/283657
;; PRIOR FILING DATE: 2001-04-13
;; PRIOR APPLICATION NUMBER: 60/283710
;; PRIOR FILING DATE: 2001-04-13
;; PRIOR APPLICATION NUMBER: 60/283678
;; PRIOR FILING DATE: 2001-04-13
;; Remaining Prior Application data removed - See file wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 251
;; SEQ ID NO 23
;; LENGTH: 1671
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (31)..(1663)
PCT-US02-10366-23
```

```
Query Match          9.3%; Score 35.2; DB 1; Length 1671;
Best Local Similarity 53.7%; Pred. No. 0.34;
Matches 73; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
```

```
OY 91 TTGAATGCTGTTCTGGAGTCGAGGATGACGCGTTGTTAGCCCTGTTCTTTCCGTT 150
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1092 TGGAACTCTACGCTCTGTTGGTTGGAGTCATCAGATACCTGGGTTATTTCCAGGC 1151
OY 151 GTTAAATCTAATGTTCTTGGAAATAAACCTCCCTGCCAAGTACTGTTGTTATGCG 210
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1152 ATATATATGACTGATTTTAAACATGACGCGCTCAGCTGCCAAAGTCTTCGGTTTGTGC 1211
OY 211 TCACATGCTTGACT 226
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1212 TTGTGCTGCTATGATT 1227
```

```
RESULT 2
PCT-US02-10366-21
; Sequence 21, Application PC/TUS0210366
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glenda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zernhusen, Bryan
; APPLICANT: Tchernen, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Verne, Corine
; APPLICANT: Spytek, Kimberly
; APPLICANT: Malyankar, Uriel
; APPLICANT: Paturajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; APPLICANT: Heyes, Melvyn
; APPLICANT: Ju, Jingtang
; APPLICANT: Peyman, John
; APPLICANT: Catterton, Elina
; APPLICANT: Macdougall, John
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Mazur, Ann
; TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES,
; TITLE OF INVENTION: NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE ANTIGENS, AND METHODS OF USE
; FILE REFERENCE: 21402-322A-061
```

```
;; CURRENT APPLICATION NUMBER: PCT/US02/10366
;; CURRENT FILING DATE: 2002-04-03
;; PRIOR APPLICATION NUMBER: 10/114153
;; PRIOR FILING DATE: 2002-04-02
;; PRIOR APPLICATION NUMBER: 60/281086
;; PRIOR FILING DATE: 2001-04-03
;; PRIOR APPLICATION NUMBER: 60/281906
;; PRIOR FILING DATE: 2001-04-05
;; PRIOR APPLICATION NUMBER: 60/282020
;; PRIOR FILING DATE: 2001-04-06
;; PRIOR APPLICATION NUMBER: 60/282930
;; PRIOR FILING DATE: 2001-04-10
;; PRIOR APPLICATION NUMBER: 60/283512
;; PRIOR FILING DATE: 2001-04-12
;; PRIOR APPLICATION NUMBER: 60/283444
;; PRIOR FILING DATE: 2001-04-12
;; PRIOR APPLICATION NUMBER: 60/283657
;; PRIOR FILING DATE: 2001-04-13
;; PRIOR APPLICATION NUMBER: 60/283710
;; PRIOR FILING DATE: 2001-04-13
;; PRIOR APPLICATION NUMBER: 60/283678
;; PRIOR FILING DATE: 2001-04-13
;; Remaining Prior Application data removed - See file wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 251
;; SEQ ID NO 21
;; LENGTH: 1677
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (25)..(1645)
PCT-US02-10366-21
```

```
Query Match          9.3%; Score 35.2; DB 1; Length 1677;
Best Local Similarity 53.7%; Pred. No. 0.34;
Matches 73; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
```

```
OY 91 TTGAATGCTGTTCTGGAGTCGAGGATGACGCGTTGTTAGCCCTGTTCTTTCCGTT 150
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1074 TGGAACTCTACGCTCTGTTGGTTGGAGTCATCAGATACCTGGGTTATTTCCAGGC 1133
OY 151 GTTAAATCTAATGTTCTTGGAAATAAACCTCCCTGCCAAGTACTGTTGTTATGCG 210
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1134 ATATATATGACTGATTTTAAACATGACGCGCTCAGCTGCCAAAGTCTTCGGTTTGTGC 1193
OY 211 TCACATGCTTGACT 226
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1194 TTGTGCTGCTATGATT 1209
```

```
RESULT 3
PCT-US02-10366-27
; Sequence 27, Application PC/TUS0210366
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glenda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zernhusen, Bryan
; APPLICANT: Tchernen, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Verne, Corine
; APPLICANT: Spytek, Kimberly
; APPLICANT: Malyankar, Uriel
; APPLICANT: Paturajan, Meera
; APPLICANT: Miller, Charles
```

```

APPLICANT: Taupier, Raymond J. Jr.
APPLICANT: Heyes, Melvyn
APPLICANT: Ju, Jinfang
APPLICANT: Peyman, John
APPLICANT: Catterton, Elina
APPLICANT: Macdougall, John
APPLICANT: Edinger, Shlomit
APPLICANT: Stone, David
APPLICANT: Mazur, Ann
TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES,
TITLE OF INVENTION: NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE ANTIGENS, AND METHODS OF USE
FILE REFERENCE: 21402-322A-061
CURRENT APPLICATION NUMBER: PCT/US02/10366
PRIOR FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: 10/114153
PRIOR FILING DATE: 2002-04-02
PRIOR APPLICATION NUMBER: 60/281086
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/281906
PRIOR FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 60/282020
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: 60/282930
PRIOR FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: 60/283512
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/283444
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/283657
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/283710
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/283678
PRIOR FILING DATE: 2001-04-13
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 251
SEQ ID NO 27
LENGTH: 2067
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (93)..(1791)
PCT-US02-10366-27

Query Match          9.3%; Score 35.2; DB 1; Length 2067;
Best Local Similarity 53.7%; Pred. No. 0.37; Mismatches 63; Indels 0; Gaps 0;
Matches 73; Conservative 0;

OY 91 TTGAATGTGTTCTGGAGCTGAAGATGACGAGTGTGTTAAGCCCTGTTCTTTCCGTT 150
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1238 TGGAACTCTACGCTCTGTTGGTTGGAGTCATCAGATACCTGGTTATTTCCAGGC 1297
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 151 GTTAAATCTAATGTTCTTGGATATAAAACCTCCCTGCCAAGTAGTACTGTTTATGC 210
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1298 ATATAATGTAAGTATTTAACAATGACGAGCCCTCACTGCCAAAGTCTTGGTTGTGC 1357
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 211 TCAACATGCTTGACT 226
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1358 TTGTCTGTGATGAT 1373
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 4
PCT-US02-10366-25
Sequence 25, Application PC/TUS0210366
GENERAL INFORMATION:
APPLICANT: Padigaru, Muralidhara
APPLICANT: Shenoy, Suresh
APPLICANT: Kekuda, Ramesh
APPLICANT: Rastelli, Luca
APPLICANT: Mezes, Peter
APPLICANT: Smithson, Glenda

```

```

APPLICANT: Guo, Xiaojia
APPLICANT: Gerlach, Valerie
APPLICANT: Casman, Stacie
APPLICANT: Boldog, Ferenc
APPLICANT: Li, Li
APPLICANT: Zernusen, Bryan
APPLICANT: Tchernev, Velizar
APPLICANT: Gangolli, Esha
APPLICANT: Vernet, Corine
APPLICANT: Spytek, Kimberly
APPLICANT: Malyankar, Uriel
APPLICANT: Pattirajan, Meera
APPLICANT: Miller, Charles
APPLICANT: Taupier, Raymond J. Jr.
APPLICANT: Heyes, Melvyn
APPLICANT: Ju, Jinfang
APPLICANT: Peyman, John
APPLICANT: Catterton, Elina
APPLICANT: Macdougall, John
APPLICANT: Edinger, Shlomit
APPLICANT: Stone, David
APPLICANT: Mazur, Ann
TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES,
TITLE OF INVENTION: NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE ANTIGENS, AND METHODS OF USE
FILE REFERENCE: 21402-322A-061
CURRENT APPLICATION NUMBER: PCT/US02/10366
PRIOR FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: 10/114153
PRIOR FILING DATE: 2002-04-02
PRIOR APPLICATION NUMBER: 60/281086
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/281906
PRIOR FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 60/282020
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: 60/282930
PRIOR FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: 60/283512
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/283444
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/283657
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/283710
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/283678
PRIOR FILING DATE: 2001-04-13
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 251
SEQ ID NO 25
LENGTH: 2130
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (31)..(1645)
PCT-US02-10366-25

Query Match          9.3%; Score 35.2; DB 1; Length 2130;
Best Local Similarity 53.7%; Pred. No. 0.37; Mismatches 63; Indels 0; Gaps 0;
Matches 73; Conservative 0;

OY 91 TTGAATGTGTTCTGGAGCTGAAGATGACGAGTGTGTTAAGCCCTGTTCTTTCCGTT 150
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1092 TGGAACTCTACGCTCTGTTGGTTGGAGTCATCAGATACCTGGTTATTTCCAGGC 1151
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 151 GTTAAATCTAATGTTCTTGGATATAAAACCTCCCTGCCAAGTAGTACTGTTTATGC 210
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1152 ATATAATGTAAGTATTTAACAATGACGAGCCCTCACTGCCAAAGTCTTGGTTGTGC 1211
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 211 TCAACATGCTTGACT 226
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```



```
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1629)
OTHER INFORMATION:
PCT-US02-38526-587

Query Match
Best Local Similarity 52.9%; Pred. No. 1.1;
Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

OY 91 TTGAATGCTGCTGCGAGCGTGAAGATGCAAGCGTTGTAAGCCCCCTGTTCCGTT 150
DB 1074 TGGAACTCTACGCTCTGTTGGGTGAGTCAATCAATACCTGGGTTATTCAGGC 1133
OY 151 GTTAACTAATGTTCTTTGGATPAAAACCTCCCTGCCAAGTACTGTTTATGCG 210
DB 1134 ATATAATGCTGATTTTAAACATGCAAGCCCTCACTGCCAAAGTTCTCGTTTGTGC 1193
OY 211 TCAACATGCTTTGACT 226
DB 1194 TTGTGCTGATGATT 1209

RESULT 8
US-09-513-999C-33882/C
Sequence 33882, Application US/09513999C
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclert, A.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
FILE REFERENCE: 59, US2, REG
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent, Pm
SEQ ID NO 33882
LENGTH: 158
TYPE: DNA
ORGANISM: Homo sapiens
US-09-513-999C-33882

Query Match
Best Local Similarity 53.4%; Pred. No. 0.6;
Matches 70; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 1 TACGGCTGCTTAAGACGACTTAAGGGGAGATGACGCGGCTCTTAGAGAAATATGG 60
DB 144 TATAGCTCTGAGAGGAGCCAAAGGAAATCTCCATTCATCTCTGCTACAGAGA 85
OY 61 AAAACACCAAGCCGAGCTCTCACAAGCTTGATGTGTCTTCTGAGACTGAAGATGC 120
DB 84 ACAGCATTAATCTAGAGGCACTAACAGCTGGAAGAGTGAATGTGCTGACTGAGA 25
OY 121 ACGTTGTTAA 131
DB 24 AGGGAGAAAA 14

RESULT 9
PCT-US02-32700-7
Sequence 7, Application PC/TUS0232700
GENERAL INFORMATION:
APPLICANT: Keith, Tim
APPLICANT: Little, Randall D.
APPLICANT: Van Berdegh, Paul
APPLICANT: Dupuis, Josee
APPLICANT: Del Mastro, Richard G.
APPLICANT: Allen, Kristina
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES
```

```
TITLE OF INVENTION: RELATING TO RESPIRATORY DISEASES AND OBESITY
FILE REFERENCE: 2976-4045
CURRENT APPLICATION NUMBER: PCT/US02/32700
CURRENT FILING DATE: 2002-10-11
PRIOR APPLICATION NUMBER: 60/328,424
PRIOR FILING DATE: 2001-10-11
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 45450
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(45450)
OTHER INFORMATION: n = A,T,C or G
PCT-US02-32700-7

Query Match
Best Local Similarity 46.6%; Pred. No. 6.7; Length 45450;
Matches 104; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

OY 133 CCCCTGTCCTTCCGTTTAACTAAATGTTCTTGGAAATPAAAACCTCCCTGCCAG 192
DB 2864 CTCATATTCCTCTGATGTTTGTATGTCCTCCCTCACTAAATPAGGCTCCCTAA 2923
OY 193 TAGTACTGTTTATGCTCAACATGCTTGACTGTTGAAGAAGACCTTGGCACAT 252
DB 2924 CAGTAGAGATCTTAATCTGATGATGTAATGAATGATGATGATGATGATGATG 2983
OY 253 TGAAGGATGCTATGAGATGCAATGCCAATCCATGGAATCAGGTGGCGGAGATGTG 312
DB 2984 TGAATAATGAATAGAGACAGTGAATGATGATGATGATGATGATGATGATGAT 3043
OY 313 CTATACAGAAATCTTCTTGCAAGATTCCTCCGGAAGA 355
DB 3044 CCAGAAATGCTACTGATCACAAGGTTCTCAGTAGAGGA 3086

RESULT 10
US-10-271-416-7
Sequence 7, Application US/10271416
GENERAL INFORMATION:
APPLICANT: Keith, Tim
APPLICANT: Little, Randall D.
APPLICANT: Van Berdegh, Paul
APPLICANT: Dupuis, Josee
APPLICANT: Del Mastro, Richard G.
APPLICANT: Allen, Kristina
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO RESPIRATORY DISEASES AND OBESITY
FILE REFERENCE: 2976-4045
CURRENT APPLICATION NUMBER: US/10/271,416
CURRENT FILING DATE: 2002-10-11
PRIOR APPLICATION NUMBER: 60/328,424
PRIOR FILING DATE: 2001-10-11
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 45450
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(45450)
OTHER INFORMATION: n = A,T,C or G
US-10-271-416-7

Query Match
Best Local Similarity 46.6%; Pred. No. 6.7; Length 45450;
Matches 104; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

OY 133 CCCCTGTCCTTCCGTTTAACTAAATGTTCTTGGAAATPAAAACCTCCCTGCCAG 192
```


Search completed: January 31, 2003, 01:19:25
Job time : 239 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2003, 01:12:19 ; Search time 3488 Seconds
(without alignments)
2724.731 Million cell updates/sec

Title: US-09-803-719-222
378
Sequence: 1 tacgcgtgcttaagacgact.....tacccattgagacacatgcg 378

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```

Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq:*

```

Result No.	Score	% Query Match	Length	ID	Description
1	378	100.0	378	US-09-803-719-222	Sequence 222, App
2	342	90.5	512	US-09-621-976-3594	Sequence 3594, App
3	342	90.5	512	US-09-834-366-8066	Sequence 8066, App
4	342	90.5	512	US-60-147-499-3594	Sequence 3594, App
5	299.6	79.3	512	US-60-197-873-8066	Sequence 8066, App
6	279.6	74.0	327	US-09-803-719-226	Sequence 226, App
7	275	72.8	312	US-09-803-719-205	Sequence 300, App
8	226.8	60.0	562	US-09-718-447-565	Sequence 565, App
9	226.8	60.0	562	US-09-534-844A-51172	Sequence 5172, App
10	131	34.7	403	US-09-689-907-861	Sequence 861, App
11	83.4	22.1	524	US-10-098-754-18036	Sequence 18036, App
12	79.4	21.0	806	US-09-685-043-1053	Sequence 1053, App
13	75.6	20.0	125	US-09-362-510-51216	Sequence 51216, App
14	48.8	12.9	381	US-09-362-510A-51216	Sequence 51216, App
15	48.8	12.9	381	US-09-904-013-51216	Sequence 28514, App
16	48.8	12.9	381	US-09-306-350A-28514	Sequence 28514, App
17	40.2	10.6	453	US-09-909-628-28514	Sequence 5950, App
18	40.2	10.6	453	US-09-652-123-5950	Sequence 14754, App
19	39.8	10.5	613	US-09-505-533-14754	Sequence 14754, App
20	39.2	10.4	1266	US-09-819-091A-14754	Sequence 14754, App
21	39.2	10.4	1266	US-09-819-091A-14754	Sequence 14754, App

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Sequence 3096, Ap	QY	1	TAGGCGCTTAAAGACACTTAAAGGGGAAATGACGACGCGCTTTAGAGAAACATATGG	60
Sequence 363, App		1	TAGGCGCTTAAAGACACTTAAAGGGGAAATGACGACGCGCTTTAGAGAAACATATGG	60
Sequence 363, App	Db	1	TAGGCGCTTAAAGACACTTAAAGGGGAAATGACGACGCGCTTTAGAGAAACATATGG	60
Sequence 14635, A				
Sequence 83432, A				
Sequence 82309, A	QY	61	AAACACCCCAAGCCGAGTCTTCACACGCTTGAATGTGTGTTGGAGCTGAAGGATGC	120
Sequence 23, Ap1	Db	61	AAACACCCCAAGCCGAGTCTTCACACGCTTGAATGTGTGTTGGAGCTGAAGGATGC	120
Sequence 23, Ap1				
Sequence 21, Ap1	QY	121	ACGGTGTATGAGCCCGCTCTTTCCGTGTATCTATGTCTTGGATATAAAC	180
Sequence 21, Ap1				
Sequence 27, Ap1	Db	121	ACGGTGTATGAGCCCGCTCTTTCCGTGTATCTATGTCTTGGATATAAAC	180
Sequence 27, Ap1				
Sequence 25, Ap1	QY	181	CTCCGTGCCAAGTAGTACTTGTTTATGTCTCAACATGCTTGACTGTGAAAAGACCC	240
Sequence 25, Ap1				
Sequence 7989, Ap	Db	181	CTCCGTGCCAAGTAGTACTTGTTTATGTCTCAACATGCTTGACTGTGAAAAGACCC	240
Sequence 256548, Sequence 256549, Sequence 256549, Sequence 30211, A	QY	241	TTTGGCACATTTGAAGGATGTGGATGAGATGCCAATCATGGATACAGTGGCGAG	300
Sequence 30211, A	Db	241	TTTGGCACATTTGAAGGATGTGGATGAGATGCCAATCATGGATACAGTGGCGAG	300
Sequence 4852, Ap	QY	301	CTATGTTGTAGTATATAGCAGAAATCTTCTTGCGCAAGATTCCTCCGGGAAAGAGGTA	360
Sequence 82461, A				
Sequence 96308, A	Db	301	CTATGTTGTAGTATATAGCAGAAATCTTCTTGCGCAAGATTCCTCCGGGAAAGAGGTA	360
Sequence 140225, A				

```

QY      301  CCATGGAGAACCATGCG  378
          |||||
Db      361  CCATGGAGAACCATGCG  378

```

RESULT 2
US-09-621-976-3594

[illegible]

```

Db 257 GGAGATGCCAATCCANCGAATACGTCGACAGCATTGTTGGTAGCTATAGACGAAGTCT 316
OY 328 TCTTGCAAAAGATTCTCCCGGGAAGAGTACCATTTGGAGACCATGTC 377
    |||||||
Db 317 TCTTGCAAAAGATTCTCCAGGGAAGAGTACCATTTGGAGAACCATGTC 366

RESULT 3
US-09-834-366-8066
; Sequence 8066, Application US/09834366
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Jobert, Severin
; APPLICANT: Giordano, Jean-Yves
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: 81 US2, REG
; CURRENT APPLICATION NUMBER: US/09/834,366
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/197,873
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 52153
; SOFTWARE: Patent.pm
; SEQ ID NO 8066
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 84..404
US-09-834-366-8066

Query Match          90.5%; Score 342; DB 32; Length 512;
Best Local Similarity 98.6%; Pred. No. 6.3e-95;
Matches 345; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 28 GAATGACGACGGGCTCTTAGAGACATATGAAACACCCAGCGAGTCTCTCACA 87
    |||||||
Db 17 GAATGACGACGACGCTCTGAGAGACATATAGAAACACCCAGCGAGTCTCTCACA 76

OY 88 AGCTGAATGTGTGTCTGAGCTGGAAGATGACACGGTGTGTAAGCCCTCTCTTCC 147
    |||||||
Db 77 AGCTGAATGTGTGTCTGAGCTGGAAGATGACACGGTGTGTAAGCCCTCTCTTCC 136

OY 148 GTTGTTAATCTAATGTCTTGGAAATPAAAAACCTCCCTGCCAAGTAGTACTTGGTTTA 207
    |||||||
Db 137 GTTGTTAATCTAATGTCTTGGAAATPAAAAACCTCCCTGCCAAGTAGTACTTGGTTTA 196

OY 208 TGCTCAACATGCTTGTGACTGTGAAAAGACCTTTGGCACACATTTGAAGGGATGTGAT 267
    |||||||
Db 197 TGCTCAACATGCTTGTGACTGTGAAAAGACCTTTGGCACACATTTGAAGGGATGTGAT 256

OY 268 GGAGATGCCAATCCATGGAATCAGGTGGCGACGATATGTTGGTACTATAGAGAAAGTCT 327
    |||||||
Db 257 GGAGATGCCAATCCATGGAATCAGGTGGCGACGATATGTTGGTACTATAGAGAAAGTCT 316

OY 328 TCTTGCAAAAGATTCTCCCGGGAAGAGTACCATTTGGAGAACCATGTC 377
    |||||||
Db 317 TCTTGCAAAAGATTCTCCAGGGAAGAGTACCATTTGGAGAACCATGTC 366

RESULT 4
US-60-147-499-3594
; Sequence 3594, Application US/60147499
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET, 054PR2
; CURRENT APPLICATION NUMBER: US/60/147,499
; CURRENT FILING DATE: 1999-08-05

```

```

; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 3594
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 84..404
US-60-147-499-3594

Query Match          90.5%; Score 342; DB 58; Length 512;
Best Local Similarity 98.6%; Pred. No. 6.3e-95;
Matches 345; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 28 GAATGACGACGGGCTCTTAGAGACATATGAAACACCCAGCGAGTCTCTCACA 87
    |||||||
Db 17 GAATGACGACGACGCTCTGAGAGACATATAGAAACACCCAGCGAGTCTCTCACA 76

OY 88 AGCTGAATGTGTGTCTGAGCTGGAAGATGACACGGTGTGTAAGCCCTCTCTTCC 147
    |||||||
Db 77 AGCTGAATGTGTGTCTGAGCTGGAAGATGACACGGTGTGTAAGCCCTCTCTTCC 136

OY 148 GTTGTTAATCTAATGTCTTGGAAATPAAAAACCTCCCTGCCAAGTAGTACTTGGTTTA 207
    |||||||
Db 137 GTTGTTAATCTAATGTCTTGGAAATPAAAAACCTCCCTGCCAAGTAGTACTTGGTTTA 196

OY 208 TGCTCAACATGCTTGTGACTGTGAAAAGACCTTTGGCACACATTTGAAGGGATGTGAT 267
    |||||||
Db 197 TGCTCAACATGCTTGTGACTGTGAAAAGACCTTTGGCACACATTTGAAGGGATGTGAT 256

OY 268 GGAGATGCCAATCCATGGAATCAGGTGGCGACGATATGTTGGTACTATAGAGAAAGTCT 327
    |||||||
Db 257 GGAGATGCCAATCCATGGAATCAGGTGGCGACGATATGTTGGTACTATAGAGAAAGTCT 316

OY 328 TCTTGCAAAAGATTCTCCCGGGAAGAGTACCATTTGGAGAACCATGTC 377
    |||||||
Db 317 TCTTGCAAAAGATTCTCCAGGGAAGAGTACCATTTGGAGAACCATGTC 366

RESULT 5
US-60-197-873-8066
; Sequence 8066, Application US/60197873
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Jobert, Severin
; APPLICANT: Giordano, Jean-Yves
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: 81 US1, PRO
; CURRENT APPLICATION NUMBER: US/60/197,873
; CURRENT FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 52153
; SOFTWARE: Patent.pm
; SEQ ID NO 8066
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 84..404
US-60-197-873-8066

Query Match          90.5%; Score 342; DB 63; Length 512;
Best Local Similarity 98.6%; Pred. No. 6.3e-95;
Matches 345; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 28 GAATGACGACGGGCTCTTAGAGACATATGAAACACCCAGCGAGTCTCTCACA 87
    |||||||
Db 17 GAATGACGACGACGCTCTGAGAGACATATAGAAACACCCAGCGAGTCTCTCACA 76

OY 88 AGCTGAATGTGTGTCTGAGCTGGAAGATGACACGGTGTGTAAGCCCTCTCTTCC 147
    |||||||

```

Db 77 AGCTTGAATGTGTCTTGGAGCTGAGAGATGACAGGTTGTAACCCCTGTTCTTTCC 136
Oy 148 GTTGTTAATCTAATGTCCTTTGGAATAAAACCTCCCTGCAATAGTACTGTTTGA 207
Db 137 GTTGTTAATCTAATGTCCTTTGGAATAAAACCTCCCTGCAATAGTACTGTTTGA 196
Oy 208 TGCTCAACATGCTTTGACTGTTGAAAAGACCTTTGGCACACATTTGAAGGATGAT 267
Db 197 TGCTCAACATGCTTTGACTGTTGAAAAGACCTTTGGCACACATTTGAAGGATGAT 256
Oy 268 GGAGATGCCAATCCATGGAATCAGGTGGCGACGCTATGTTGGTATATAGCAGAACT 327
Db 257 GGAGATGCCAATCCATGGAATCAGGTGGCGACGCTATGTTGGTATATAGCAGAACT 316
Oy 328 TCTTGGCAAAAGATTCCTCCCGGAAGGAAGTACCATGAGCAACCATGC 377
Db 317 TCTTGGCAAAAGATTCCTCCCGGAAGGAAGTACCATGAGCAACCATGC 366

RESULT 6

US-09-803-719-276
Sequence 276, Application US/09803719
GENERAL INFORMATION:
APPLICANT: Williams, Lewis T.
APPLICANT: Escobedo, Jaime
APPLICANT: Innis, Michael A.
APPLICANT: Garcia, Pablo Dominguez
APPLICANT: Sudduth-Klinger, Julie
APPLICANT: Reinhard, Christoph
APPLICANT: Glese, Klaus
APPLICANT: Randazzo, Filippo
APPLICANT: Kennedy, Giulia C.
APPLICANT: Pot, David
APPLICANT: Kassam, Altaf
APPLICANT: Lamson, George
APPLICANT: Drmanac, Radoje
APPLICANT: Cirvenjakov, Radomir
APPLICANT: Dickson, Mark
APPLICANT: Leshkowitz, Dena
APPLICANT: Labat, Ivan
APPLICANT: Leshkowitz, Dena
APPLICANT: Kita, David
APPLICANT: Garcia, Veronica
APPLICANT: Jones, Lee William
APPLICANT: Stache-Crain, Birgit
TITLE OF INVENTION: Human Genes and Gene Products
FILE REFERENCE: 1624.002
CURRENT APPLICATION NUMBER: US/09/803,719
CURRENT FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/188,609
PRIOR FILING DATE: 2000-03-09
NUMBER OF SEQ ID NOS: 2396
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 276
LENGTH: 327
TYPE: DNA
ORGANISM: Homo sapiens
US-09-803-719-276

Query Match 79.3%; Score 299.6; DB 31; Length 327;
Best Local Similarity 98.7%; Pred. No. 7.5e-82;
Matches 302; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 28 GAATGACGACGCGCTCTTAGAGAACTATGGAACACCCAGCCGAGTCTCTCACA 87
Db 21 GAATGACGACGCGCTCTTAGAGAACTATGGAACACCCAGCCGAGTCTCTCACA 80
Oy 88 AGCTTGAATGTGTCTTGGAGCTGAAGATGACAGGTTGTTAAAGCCCTGTTCTTTCC 147
Db 81 AGCTTGAATGTGTCTTGGAGCTGAAGATGACAGGTTGTTAAAGCCCTGTTCTTTCC 140
Oy 148 GTTGTTAATCTAATGTCCTTTGGAATAAAACCTCCCTGCAATAGTACTGTTTGA 207

Db 141 GTTGTTAATCTAATGTCCTTTGGAATAAAACCTCCCTGCAATAGTACTGTTTGA 200
Oy 208 TGCTCAACATGCTTTGACTGTTGAAAAGACCTTTGGCACACATTTGAAGGATGAT 267
Db 201 TGCTCAACATGCTTTGACTGTTGAAAAGACCTTTGGCACACATTTGAAGGATGAT 260
Oy 268 GGAGATGCCAATCCATGGAATCAGGTGGCGACGCTATGTTGGTATATAGCAGAACT 327
Db 261 GGAGATGCCAATCCATGGAATCAGGTGGCGACGCTATGTTGGTATATAGCAGAACT 320
Oy 328 TCTTGG 333
Db 321 TCTTGG 326

RESULT 7

US-09-803-719-295
Sequence 295, Application US/09803719
GENERAL INFORMATION:
APPLICANT: Williams, Lewis T.
APPLICANT: Escobedo, Jaime
APPLICANT: Innis, Michael A.
APPLICANT: Garcia, Pablo Dominguez
APPLICANT: Sudduth-Klinger, Julie
APPLICANT: Reinhard, Christoph
APPLICANT: Glese, Klaus
APPLICANT: Randazzo, Filippo
APPLICANT: Kennedy, Giulia C.
APPLICANT: Pot, David
APPLICANT: Kassam, Altaf
APPLICANT: Lamson, George
APPLICANT: Drmanac, Radoje
APPLICANT: Cirvenjakov, Radomir
APPLICANT: Dickson, Mark
APPLICANT: Leshkowitz, Dena
APPLICANT: Labat, Ivan
APPLICANT: Leshkowitz, Dena
APPLICANT: Kita, David
APPLICANT: Garcia, Veronica
APPLICANT: Jones, Lee William
APPLICANT: Stache-Crain, Birgit
TITLE OF INVENTION: Human Genes and Gene Products
FILE REFERENCE: 1624.002
CURRENT APPLICATION NUMBER: US/09/803,719
CURRENT FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/188,609
PRIOR FILING DATE: 2000-03-09
NUMBER OF SEQ ID NOS: 2396
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 295
LENGTH: 322
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)-(322)
OTHER INFORMATION: n - A,T,C or G
US-09-803-719-295

Query Match 74.0%; Score 279.6; DB 31; Length 322;
Best Local Similarity 98.6%; Pred. No. 1.2e-75;
Matches 282; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 28 GAATGACGACGCGCTCTTAGAGAACTATGGAACACCCAGCCGAGTCTCTCACA 87
Db 36 GAATGACGACGCGCTCTTAGAGAACTATGGAACACCCAGCCGAGTCTCTCACA 95
Oy 88 AGCTTGAATGTGTCTTGGAGCTGAAGATGACAGGTTGTTAAAGCCCTGTTCTTTCC 147
Db 96 AGCTTGAATGTGTCTTGGAGCTGAAGATGACAGGTTGTTAAAGCCCTGTTCTTTCC 155
Oy 148 GTTGTTAATCTAATGTCCTTTGGAATAAAACCTCCCTGCAATAGTACTGTTTGA 207

```
Db 156 GTTGTATTAATCTAATGTTCTTTGGAATAAAACCTCCCTGCCAAGTAGTACTGTTTGA 215
Oy 208 TGCCTCAACATGCTTTGACCTGTGAAAAGAGACCTTTGGCAGACATTTGAAGGATGTTGAT 267
Db 216 TGCCTCAACATGCTTTGACCTGTGAAAAGAGACCTTTGGCAGACATTTGAAGGATGTTGAT 275
Oy 268 GGAGATGCCAATCCATCGAATCAGGTGGCGCAGCTATGTTGGTAGC 313
Db 276 GGAGATGCCAATCCATCGAATCAGGTGGCGCAGCTATGTTGGTAGC 321

RESULT 8
US-09-803-719-300
Sequence 300, Application US/09803719
GENERAL INFORMATION:
APPLICANT: Williams, Lewis T.
APPLICANT: Escobedo, Jaime
APPLICANT: Innis, Michael A.
APPLICANT: Garcia, Pablo Dominguez
APPLICANT: Suduth-Klinger, Julie
APPLICANT: Reinhard, Christoph
APPLICANT: Glese, Klaus
APPLICANT: Randazzo, Filippo
APPLICANT: Kennedy, Giulia C.
APPLICANT: Pot, David
APPLICANT: Kaasam, Altaf
APPLICANT: Lamson, George
APPLICANT: Drmanac, Radoje
APPLICANT: Ckrvenjakov, Radomir
APPLICANT: Dickson, Mark
APPLICANT: Drmanac, Snezana
APPLICANT: Labat, Ivan
APPLICANT: Leashkowitz, Dena
APPLICANT: Kita, Veronica
APPLICANT: Garcia, Lee William
APPLICANT: Jones, Lee William
APPLICANT: Stache-Crain, Birgit
TITLE OF INVENTION: Human Genes and Gene Products
FILE REFERENCE: 1624.002
CURRENT APPLICATION NUMBER: US/09/803,719
CURRENT FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/188,609
PRIOR FILING DATE: 2000-03-09
NUMBER OF SEQ ID NOS: 2396
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 300
LENGTH: 318
TYPE: DNA
ORGANISM: Homo sapiens
US-09-803-719-300

Query Match 72.8%; Score 275; DB 31; Length 318;
Best Local Similarity 98.2%; Pred. No. 3.2e-74;
Matches 278; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
Db 276 GGAGATGCCAATCCATCGAATCAGGTGGCGCAGCTATGTTGGTAGC 318

RESULT 9
US-09-718-457-565
Sequence 565, Application US/09718457
GENERAL INFORMATION:
APPLICANT: Watlier, Sigrid
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
TITLE OF INVENTION: Novel Human Polynucleotides and
FILE REFERENCE: LEX-0093-USA
CURRENT APPLICATION NUMBER: US/09/718,457
CURRENT FILING DATE: 2000-11-22
PRIOR APPLICATION NUMBER: US 60/167,870
PRIOR FILING DATE: 1999-11-29
NUMBER OF SEQ ID NOS: 1008
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 565
LENGTH: 562
TYPE: DNA
ORGANISM: Homo sapiens
US-09-718-457-565

Query Match 60.0%; Score 226.8; DB 28; Length 562;
Best Local Similarity 99.1%; Pred. No. 3.5e-59;
Matches 228; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 148 GTTGTATTAATCTAATGTTCTTTGGAATAAAACCTCCCTGCCAAGTAGTACTGTTTGA 207
Db 1 GTTGTATTAATCTAATGTTCTTTGGAATAAAACCTCCCTGCCAAGTAGTACTGTTTGA 60
Oy 208 TGCCTCAACATGCTTTGACCTGTGAAAAGAGACCTTTGGCAGACATTTGAAGGATGTTGAT 267
Db 61 TGCCTCAACATGCTTTGACCTGTGAAAAGAGACCTTTGGCAGACATTTGAAGGATGTTGAT 120
Oy 268 GGAGATGCCAATCCATCGAATCAGGTGGCGCAGCTATGTTGGTAGTATAGCAGAGTCT 327
Db 121 GGAGATGCCAATCCATCGAATCAGGTGGCGCAGCTATGTTGGTAGTATAGCAGAGTCT 180
Oy 328 TCTTGGCAAGATTTCTCCCGGGAAGAGAGGTACCATTTGGAAACCATGC 377
Db 181 TCTTGGCAAGATTTCTCCCGGGAAGAGAGGTACCATTTGGAAACCATGC 230

RESULT 10
US-09-534-844A-5172
Sequence 5172, Application US/09534844A
GENERAL INFORMATION:
APPLICANT: Drmanac, Radoje T.
APPLICANT: Labat, Ivan
APPLICANT: Stache-Crain, Birgit
APPLICANT: Dickson, Mark C.
APPLICANT: Jones, Lee W.
TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
FILE REFERENCE: 21272-109 (775)
CURRENT APPLICATION NUMBER: US/09/534,844A
CURRENT FILING DATE: 2001-03-24
PRIOR APPLICATION NUMBER: 60/126,605
PRIOR FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 14510
SOFTWARE: Hy-patent.pl Version 1.1
SEQ ID NO 5172
LENGTH: 403
TYPE: DNA
ORGANISM: Homo sapiens
US-09-534-844A-5172

Query Match 34.7%; Score 131; DB 20; Length 403;
Best Local Similarity 95.7%; Pred. No. 1.6e-29;
```

Matches	156	Conservative	0	Mismatches	5	Indels	2	Gaps	2
OY	148	GTGTTTAACTAACTATGTTCTTTGGAAATAAAACCTCCCTGCCAAGTAAGTACTTGGTTTA	207						
Db	242	GCTGTTTAACTCAAAAGTTCTTTGGAAATAAAACCTCCCTGCCAAGAGACTT-GATTTA	300						
OY	208	TGCTCAACATGCTTTGACGTGGAAAGACCTTTGGCACATTTGAAGGGATGGTAT	267						
Db	301	TGCTCAACATGCTTTGACGTGGAAAGAGACCTTTGGCACATTTGAAGGGAT-GTAT	359						
OY	268	GGAGATGCCAATCCATGGAATCAGGTGGCGCAGCTATGTTGGT	310						
Db	360	GGAGATGCCAATCCATGGAATCAATGGCACACGATGTTGGT	402						

```

RESULT 11
US-09-689-907-861/c
: Sequence 861. Application US/09689907
: GENERAL INFORMATION:
: APPLICANT: Nehls, Michael
: APPLICANT: Zambrowicz, Brian
: APPLICANT: Sands, Arthur T.
: TITLE OF INVENTION: Novel Human Polynucleotides and the
: FILE REFERENCE: LEX-0064-USA
: CURRENT APPLICATION NUMBER: US/09/689,907
: CURRENT FILING DATE: 2000-10-11
: PRIOR APPLICATION NUMBER: US 60/158,799
: PRIOR FILING DATE: 1999-10-12
: NUMBER OF SEQ ID NOS: 1508
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 861
: LENGTH: 524
: TYPE: DNA
: ORGANISM: homo sapiens
US-09-689-907-861

```

	Query Match	Similarity	Score	DB	length
Best Local Similarity	98.8%	Pred. No. 1e-14;			
Matches	84;	Conservative	0;	Mismatches	1;
				Indels	0;
				Gaps	0
Oy	262	GGTGATGAGATGCCAAATCCATGATGATCAGTGGGCGACGCTATGTTGTGTAGCTATAGCAG	321		
Db	249	GGTGTGTGAGATGCCAAATCCATGATGATCAGTGGGCGACGCTATGTTGTGTAGCTATAGCAG	190		
Oy	322	AACTCTTCTTGGCAAGATTCCTCC	346		
Db	189	AACTCTTCTTGGCAAGATTCCTCC	165		

```

RESULT 12
US-10-098-754-19036/c
; Sequence 19036, Application US/10098754
; GENERAL INFORMATION:
; APPLICANT: Harrington, John J.
; APPLICANT: Jackson, P. David
; APPLICANT: Sherf, Bruce A.
; APPLICANT: Cain, Scott
; APPLICANT: Rundlett, Stephen E.
; APPLICANT: Ramachandran, Rakesh
; TITLE OF INVENTION: NOVEL NUCLEIC ACID AND PROTEIN HOMOLOGS AND USES THEREOF
; FILE REFERENCE: ATN-001CN
; CURRENT APPLICATION NUMBER: US/10/098,754
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 09/791539
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 21107
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19036
; LENGTH: 806
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: (1)...(806)
; OTHER INFORMATION: n - A,T,C or G
US-10-098-754-19036

```

Query Match	21.0%;	Score 79.4;	DB 39;	Length 806;
Best Local Similarity	-98.8%;	Pred. No. 2.2e-13;		
Matches 80;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

Oy	252	GGTATGAGAAATGCCAAATCCATGGAATCAGGTGGCGCAGCATTTGGTAGTATACAG	321
	423	GGTATGGAATGCCAATCCATGGAATCAGGTGGCACAGCATTTGGTAGTATACAG	364
Oy	322	AAGCTTCTTGGCAAAAGTTC	342
	363	AAGTTCTTGGCAAAAGTTC	343

RESULT 13
US-09-685-045-1053
; sequence 1053, Application US/09685045
; contact information

```

? APPLICANT: Neils, Michael
? APPLICANT: Zambrowicz, Brian
? APPLICANT: Sands, Arthur T.
? TITLE OF INVENTION: Novel Human Polynucleotides and
? FILE REFERENCE: LEX-0055-USA
? CURRENT APPLICATION NUMBER: US/09/685,045
? CURRENT FILING DATE: 2000-10-06
? PRIOR APPLICATION NUMBER: US 60/158,741
? PRIOR FILING DATE: 1999-10-12
? NUMBER OF SEQ ID NOS: 1508
? SOFTWARE: PasteSeq for Windows Version 4.0
? SEQ ID NO 1053
? LENGTH: 125
? TYPE: DNA
? ORGANISM: homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(125)
? OTHER INFORMATION: n = A,T,C or G
? US-09-685-045-1053

```

Query Match	20.0%;	Score 75.6;	DB 27;	Length 125;
Best Local Similarity	83.1%;	Pred. No. 1.5e-12;		
Matches 103; Conservative	0;	Mismatches 19;	Indels 2;	Gaps 2;

Accession	Sequence	Length
QY 198	CTTGGTTTATGACCAACATGCTTTGACTG-TAAAAAGACCTTTGGCACATTTGAA	256
Db 1	CTTGGTTTATGACCAACATGCTTTGACTG-TAAAAAGACCTTTGGCACATTTGAA	60
QY 257	GGGATGGGATGAGAGATGCCAATCCATGATGATCAGGTGGCCGACCTTGTTGGTAGCTAT	316
Db 61	GGATGGGATGAGATGAGATGCCAATCCATGATGATCAGGTGGCCGACCTTGTTGGTAGCTAT	119
QY 317	AGCA 320	
Db 120	AGCA 123	

RESULT 14
 US-09-362-510-51216
 Sequence 51216, Application US/09362510
 GENERAL INFORMATION:
 APPLICANT: Hyseq, Inc.
 TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 FROM VARIOUS CDNA LIBRARIES
 FILE REFERENCE: 20411-756CON1
 CURRENT APPLICATION NUMBER: US/09/362,510
 CURRENT FILING DATE: 1999-07-27
 EARLIER APPLICATION NUMBER: US 09/721,820
 EARLIER FILING DATE: 1998-12-30

NUMBER OF SEQ ID NOS: 62165
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 51216
 LENGTH: 381
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (1)...(381)
 OTHER INFORMATION: n = A,T,C or G
 US-09-362-510-51216

Query Match 12.9%; Score 48.8; DB 17; Length 381;
 Best Local Similarity 96.2%; Pred. No. 0.00049;
 Matches 50; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 326 CTTCTGGCAAGATTCTCCCGGAAGAGTACCATTTGGAGAACCATGC 377
 DB 1 CTTCTGGCAAGATTCTCCCGGAAGAGTACCATTTGGAGAACCATGC 52

RESULT 15

US-09-362-510A-51216
 Sequence 51216, Application US/09362510A
 GENERAL INFORMATION:
 APPLICANT: Hysed, Inc.
 TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 FILE REFERENCE: 20411-759CON1
 CURRENT APPLICATION NUMBER: US/09/362,510A
 CURRENT FILING DATE: 1999-07-27
 PRIOR APPLICATION NUMBER: US 09/221,820
 PRIOR FILING DATE: 1998-12-30
 NUMBER OF SEQ ID NOS: 62165
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 51216
 LENGTH: 381
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (1)...(381)
 OTHER INFORMATION: n = A,T,C or G
 US-09-362-510A-51216

Query Match 12.9%; Score 48.8; DB 17; Length 381;
 Best Local Similarity 96.2%; Pred. No. 0.00049;
 Matches 50; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 326 CTTCTGGCAAGATTCTCCCGGAAGAGTACCATTTGGAGAACCATGC 377
 DB 1 CTTCTGGCAAGATTCTCCCGGAAGAGTACCATTTGGAGAACCATGC 52

Search completed: January 31, 2003, 02:17:22
 Job time : 3495 secs

RESULT 2

US-09-737-178-103

; Sequence 103, Application US/09737178

; Patent No. US20010029295A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; APPLICANT: Sleath, Paul R.

; APPLICANT: McNeill, Patricia D.

; APPLICANT: Homer, Mary

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS

; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION

; FILE REFERENCE: 210121.426C9

; CURRENT APPLICATION NUMBER: US/09/737,178

; CURRENT FILING DATE: 2000-12-13

; NUMBER OF SEQ ID NOS: 144

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 103

; LENGTH: 2966

; TYPE: DNA

; ORGANISM: Babesia microti

US-09-737-178-103:

Query Match 8.4%; Score 31.6; DB 10; Length 2966;

Best Local Similarity 51.4%; Pred. No. 4.4;

Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 142 TTTCCGTTGTTAATCTAATGTTCTTTGGAAATAAAACCTCCCTGCCAGTAGTACTTG 201

DB 271 TATTCCTACTTTTAATATATATTTTATTCATAATATCTCTTATATTTTGTACTAT 330

QY 202 GTTTATGCTCAACAGCTTTGACTGTGAAAGAGACCTTGGCACACATTGAAGGAT 261

DB 331 TTTATATATCAT 390

QY 262 GGTGATGAGATGCCAATCCAT 283

DB 391 GATGATATACATGAAATATGCAT 412

RESULT 3

US-09-737-178-112

; Sequence 112, Application US/09737178

; Patent No. US20010029295A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; APPLICANT: Sleath, Paul R.

; APPLICANT: McNeill, Patricia D.

; APPLICANT: Homer, Mary

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS

; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION

; FILE REFERENCE: 210121.426C9

; CURRENT APPLICATION NUMBER: US/09/737,178

; CURRENT FILING DATE: 2000-12-13

; NUMBER OF SEQ ID NOS: 144

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 112

; LENGTH: 3011

; TYPE: DNA

; ORGANISM: Babesia microti

US-09-737-178-112

Query Match 8.4%; Score 31.6; DB 10; Length 3011;

Best Local Similarity 51.4%; Pred. No. 4.4;

Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 142 TTTCCGTTGTTAATCTAATGTTCTTTGGAAATAAAACCTCCCTGCCAGTAGTACTTG 201
DB 271 TATTCCTACTTTTAATATATATTTTATTCATAATATCTCTTATATTTTGTACTAT 330
QY 202 GTTTATGCTCAACAGCTTTGACTGTGAAAGAGACCTTGGCACACATTGAAGGAT 261
DB 331 TTTATATATCAT 390
QY 262 GGTGATGAGATGCCAATCCAT 283
DB 391 GATGATATACATGAAATATGCAT 412

RESULT 4

US-09-286-488-10

; Sequence 10, Application US/09286488

; Patent No. US20020169136A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; APPLICANT: Sleath, Paul R.

; APPLICANT: McNeill, Patricia D.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS

; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION

; FILE REFERENCE: 210121.426C3

; CURRENT APPLICATION NUMBER: US/09/286,488

; CURRENT FILING DATE: 1999-04-05

; NUMBER OF SEQ ID NOS: 83

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 10

; LENGTH: 3701

; TYPE: DNA

; ORGANISM: Babesia microti

US-09-286-488-10

Query Match 8.4%; Score 31.6; DB 9; Length 3701;

Best Local Similarity 51.4%; Pred. No. 4.8;

Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 142 TTTCCGTTGTTAATCTAATGTTCTTTGGAAATAAAACCTCCCTGCCAGTAGTACTTG 201
DB 999 TATTCCTACTTTTAATATATATTTTATTCATAATATCTCTTATATTTTGTACTAT 1058
QY 202 GTTTATGCTCAACAGCTTTGACTGTGAAAGAGACCTTGGCACACATTGAAGGAT 261
DB 1059 TTTATATATCAT 1118
QY 262 GGTGATGAGATGCCAATCCAT 283
DB 1119 GATGATATACATGAAATATGCAT 1140

RESULT 5

US-09-737-178-10

; Sequence 10, Application US/09737178

; Patent No. US20010029295A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; APPLICANT: Sleath, Paul R.

; APPLICANT: McNeill, Patricia D.

; APPLICANT: Homer, Mary

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS

; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION

; FILE REFERENCE: 210121.426C9

; CURRENT APPLICATION NUMBER: US/09/737,178

; CURRENT FILING DATE: 2000-12-13

; NUMBER OF SEQ ID NOS: 144

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 10

```

; LENGTH: 3701
; TYPE: DNA
; ORGANISM: Babesia microti
US-09-737-178-10

Query Match
Best Local Similarity 51.4%; Score 31.6; DB 10; Length 3701;
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 142 TTTTCGCTTTTAATGTAATGTTCTTGGATAAAGACCTCCCTCCAGTAGTACTTG 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 999 TATTCCTACTTTTAATGTAATGTTCTTGGATAAAGACCTCCCTCCAGTAGTACTTG 1058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 202 GTTTTATGCACACATCTTGTGACTGTGAAAAGACCTTGGCACACATTGAAGGAT 261
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1059 TTTTATTAATCAATATATATATATATATATATATATATATATATATATATATATAT 1118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 262 GGTGATGAGATGCCATCCAT 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1119 GATGATATACATGAAATATGCAT 1140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
US-09-864-761-12790/C
; Sequence 12790, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
```

```

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO: 12790
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011260.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.81
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.94
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.99
US-09-864-761-12790

Query Match
Best Local Similarity 64.4%; Score 31.4; DB 10; Length 568;
Matches 47; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 93 GAATGTGCTGCTGAGCTGAAGATGACAGCTTGTAAAGCCCTGTTCCGTTGT 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 272 GAATGTGCTGCTGAGCTGAAGATGAGGAAGCTTGTAAATATCATAGCTCTACTAGT 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 153 TTAATCTAATGTT 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 212 CTAATGTAATAGTT 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-10-002-344A-49/C
; Sequence 49, Application US/10002344A
; Patent No. US20020172959A1
; GENERAL INFORMATION:
; APPLICANT: Reclapon, Hervé
; APPLICANT: Sun, Yongming
; APPLICANT: Chen, Sel-Yu
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes an
; FILE REFERENCE: DEX-0241
; CURRENT APPLICATION NUMBER: US/10/002,344A
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/242,998
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 49
; LENGTH: 631
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-002-344A-49

Query Match
Best Local Similarity 63.5%; Score 30.8; DB 9; Length 631;
Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 106 GGAGCTGAAGAGACGCTGTTAGCCCTGTTCTTTCCGTTTAAATCTATGTT 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 623 GCACTGAAGGCTCAACATTATGAGCCCTCTTTTGGCATATCTTATATGATG 564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 CTTTGAATATAAA 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 563 CTTTAAATTAAAAA 550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
US-09-796-692-4354/C
; Sequence 4354, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
```

;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
;; FILE REFERENCE: 2077.001200
;; CURRENT APPLICATION NUMBER: US/09/796,692
;; CURRENT FILING DATE: 2001-03-01
;; PRIOR APPLICATION NUMBER: 60/186,126
;; PRIOR FILING DATE: 2000-03-01
;; PRIOR APPLICATION NUMBER: 60/190,479
;; PRIOR FILING DATE: 2000-03-17
;; PRIOR APPLICATION NUMBER: 60/200,545
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: 60/200,303
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: 60/200,779
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: 60/200,999
;; PRIOR FILING DATE: 2000-05-01
;; PRIOR APPLICATION NUMBER: 60/202,084
;; PRIOR FILING DATE: 2000-05-04
;; PRIOR APPLICATION NUMBER: 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: 60/223,416
;; PRIOR FILING DATE: 2000-08-04
;; PRIOR APPLICATION NUMBER: 60/223,378
;; PRIOR FILING DATE: 2000-08-07
;; NUMBER OF SEQ ID NOS: 9597
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 4354
;; LENGTH: 440
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-796-692-4354

Query Match
Best Local Similarity 47.2%; Pred. No. 3.9;
Matches 93; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 114 AGAATGACAGCTGTGTAAGCCCGCTTCTTCCGTTGTTAACTAATGTTCTTTGGAA 173
DB 395 ATGCTGCGTGTGTCAGACGCCCTCTGTGTGTGTGATTTGTTTATTTTCTTTT 336
QY 174 TAAAAACCTCCCTGCCAAGTAGTACTGTTTATGCTCAACATGCTTTGACTGTGAA 233
DB 335 TTACATATATATGACGAGGAGTAGTACTGTAATGTTTCTTATGTTGTTCA 276
QY 234 AGAGACCTTTGGACACATTGAAAGGATGATGAGATGCCAATCATGGAATCAGCT 293
DB 275 ATATGAAATTTGGAACACCAAGCCGCTAATGAGATGACAGCTTTTCTGGAACCCAGA 216
QY 294 GGGCAGCTATGTTGCT 310
DB 215 GTCAACAACCAATTGAT 199

RESULT 9
US-09-745-763-204
; Sequence 204, Application US/09745763
; Patent No. US20020065394A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; Lavalite, Edward R.
; Collins-Racie, Lisa A.
; Evans, Cheryl
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; ENCODING THEM

;; NUMBER OF SEQUENCES: 219
;; CORRESPONDENCE ADDRESS:
;; ADDRESS: Genetics Institute, Inc.
;; STREET: 87 Cambridgepark Drive
;; CITY: Cambridge
;; STATE: MA
;; COUNTRY: U.S.A.
;; ZIP: 02140
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/745,763
;; FILING DATE: 18-Jun-2000
;; CLASSIFICATION: <Unknown>
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Sprunger, Suzanne A.
;; REGISTRATION NUMBER: 41,323
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 498-8284
;; TELEFAX: (617) 876-5851
;; INFORMATION FOR SEQ ID NO: 204:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4285 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; SEQUENCE DESCRIPTION: SEQ ID NO: 204:
US-09-745-763-204

Query Match
Best Local Similarity 47.2%; Pred. No. 11;
Matches 93; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 114 AGAATGACAGCTGTGTAAGCCCGCTTCTTCCGTTGTTAACTAATGTTCTTTGGAA 173
DB 2938 ATGCTGCGTGTGTCAGACGCCCTCTGTGTGTGTGATTTGTTTATTTTCTTTT 2997
QY 174 TAAAAACCTCCCTGCCAAGTAGTACTGTTTATGCTCAACATGCTTTGACTGTGAA 233
DB 2998 TTACATATATATGACGAGGAGTAGTACTGTAATGTTTCTTATGTTGTTCA 3057
QY 234 AGAGACCTTTGGACACATTGAAAGGATGATGAGATGCCAATCATGGAATCAGCT 293
DB 3058 ATATGAAATTTGGAACACCAAGCCGCTAATGAGATGACAGCTTTTCTGGAACCCAGA 3117
QY 294 GGGCAGCTATGTTGCT 310
DB 3118 GTCAACAACCAATTGAT 3134

RESULT 10
US-09-764-877-2678/c
; Sequence 2678, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2678
; LENGTH: 26591
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2678

```

Query Match
Best Local Similarity 8.1%; Score 30.6; DB 10; Length 26591;
Best Local Similarity 58.1%; Pred. No. 24;
Matches 54; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 145 TCCGTTGTTTAATCAATGTTCTTTGGAATATAAAACCTCCCTGCCAAGTACTAGTGGTT 204
DB 15573 TCCGTTGGAATATCAATGTTCTTTCCCTGCGAGAAATGTTCTTTTAATATTTTAATGCTT 15514

OY 205 TTATGCTCAACATGCTTTGACTGTTGAAAAGAG 237
DB 15513 TAATTTTCACGCTGCTTACTGATCAAAATAG 15481

RESULT 11
US-09-822-849A-246
; Sequence 246, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukola, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 246
; LENGTH: 2648
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-246

Query Match
Best Local Similarity 8.0%; Score 30.4; DB 10; Length 2648;
Best Local Similarity 51.5%; Pred. No. 10;
Matches 70; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

OY 88 ACCTGAATGCTGTGCTTGAGCTGACGATGACGAGGTTGTTAAGCCCTGTTCTTTCC 147
DB 1060 AACTTAATGCTCAAAAGGACGACGCGGTGCGCAGTAAATTAATTAATGTTCAAC 1119

OY 148 GTTGTATTAATCAATGTTCTTTGGAATATAAAACCTCCCTGCCAAGTACTAGTGGTTT 207
DB 1120 CTATGATATAAATATACATTAATGATATAAATACGCTATATAATAGTACTTCAGGCA 1179

OY 208 TGCTCAACATGCTTTG 223
DB 1180 TTTAAACATGTTGTG 1195

RESULT 12
US-09-764-847-1518/c
; Sequence 1518, Application US/09764847
; Patent No. US20020132761A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1518
; LENGTH: 1655

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1518

Query Match
Best Local Similarity 7.9%; Score 30; DB 10; Length 1655;
Best Local Similarity 54.5%; Pred. No. 11;
Matches 60; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 80 CTCACAAAGCTGAATGTTGTTCTGAGCTGAAGATGACAGGTTGTTAAGCCCTGT 139
DB 1221 CTCACAAATATAAATATGATCCAGGCTCAAGAGGAACTGATTTTGTCTTTT 1162

OY 140 TCTTTCCGTTGTTTAATCTAATGTTCTTTGGAATATAAAACCTCCCTGCC 189
DB 1161 TGTTTCCCTTTGTTTCCCTTAATTTCCCTTGAAGAACATGTCACCTGCTC 1112

RESULT 13
US-09-952-360-1/c
; Sequence 1, Application US/09952360
; Patent No. US20020072066A1
; GENERAL INFORMATION:
; APPLICANT: Styrkarsdottir, Unnur
; APPLICANT: Johannsdottir, Vala Drott
; TITLE OF INVENTION: Human Osteoporosis Gene
; FILE REFERENCE: 2345, 2009-001
; CURRENT APPLICATION NUMBER: US/09/952,360
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: 09/661,887
; PRIOR FILING DATE: 2000-09-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 14759
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3639)...(3984)
; NAME/KEY: CDS
; LOCATION: (11757)...(12601)
US-09-952-360-1

Query Match
Best Local Similarity 7.9%; Score 30; DB 10; Length 14759;
Best Local Similarity 49.4%; Pred. No. 28;
Matches 78; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

OY 29 AATGACGACGCGCTCTTAGAGAAACATATGGAAGAAACCCAGCGAGCTCTACAA 88
DB 8179 AATGAAACAGTTGACATTCAGAGAGATATGTAACCTTCTTAAGCTACAGAGTAA 8120

OY 89 GCTTAATGTTGTTCTGAGCTGAAGATGACGAGGTTGTTAAGCCCTGTTCTTTCCG 148
DB 8119 ATTAAGTAGATATATATATATATGATCTACTGCGCTCTCAAGGTTTCTTGCTTTGTT 8060

OY 149 TTGTTTAATCTAATGTTCTTTGGAATATAAAACCTCCCT 186
DB 8059 TTGTTTGTGTTCTGTTTGTGTTTGTGACACAGGCTCACT 8022

RESULT 14
US-09-818-512-3/c
; Sequence 3, Application US/09818512
; Patent No. US20020142416A1
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEROOF
; FILE REFERENCE: C1001192
; CURRENT APPLICATION NUMBER: US/09/818,512
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 4

```

```

OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (395)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (404)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (407)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (422)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (429)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-12445

Query Match      7.88; Score 29.6; DB 10; Length 438;
Best Local Similarity 61.08; Pred. No. 8.1;
Matches 47; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY    93 GAATGTGTTCTCGAGCTGAAGATGCACGGTGTGTAAGCCCCNGTCTTTCCGTTGT 152
       ||||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    150 GAATGTGTTCTCGACGCTGATGATGNAAATGTTTGTAATAATCTCTGTAGATCATTTA 209

QY    153 TTAATCTAATGTTCCTT 169
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    210 TTATTTATTTGTTTGT 226

Search completed: January 31, 2003, 01:17:11
Job time : 157 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2003, 23:50:38 ; Search time 52 Seconds

(Without alignments)
2229,304 Million cell updates/sec

Title: US-09-803-719-222

Perfect score: 378
Sequence: 1 tacggcgtcctaagacgact.....taccattggagacaccatcg 378

Scoring table: IDENTITY-NUC
Gapop 10.0 ; Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents,NA:*
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PTCUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31.6	8.4	3701	4	US-08-845-258-10
2	31.6	8.4	3701	4	US-08-990-571-10
3	31.6	8.4	3701	4	US-08-723-142A-10
4	31.6	8.4	3701	4	US-09-528-784A-10
5	30.4	8.0	20598	4	US-09-593-995-10
6	30.4	7.9	7003	2	US-08-378-548-24
7	29.6	7.8	900	4	US-09-134-618-3
8	29.6	7.8	1392	4	US-09-006-353A-1
9	29.6	7.8	1392	4	US-09-573-986-1
10	29.6	7.8	4000	2	US-08-687-289A-2
11	29.4	7.8	3834	3	US-09-209-668-18
12	29.4	7.8	3834	1	US-08-365-470-1
13	29.4	7.8	3858	2	US-08-344-155C-98
14	29.4	7.8	3858	4	US-09-009-490A-88
15	29.4	7.8	3863	4	US-08-482-073-1
16	29.4	7.8	3863	6	5217870-1
17	29.2	7.7	864	1	US-08-204-675-3
18	29.2	7.7	864	2	US-08-660-754-3
19	29.2	7.7	864	2	US-08-796-364-3
20	29.2	7.7	864	5	PCT-US95-02520-3
21	29.2	7.7	4933	1	US-08-204-675-2
22	29.2	7.7	4933	2	US-08-796-364-2
23	29.2	7.7	4933	2	US-08-660-754-2
24	29.2	7.7	4933	2	US-08-796-364-2
25	29.2	7.7	4933	2	PCT-US95-02520-2
26	28.8	7.6	420	1	US-09-345-882-1
27	28.8	7.6	420	1	US-08-504-718D-13
27	28.8	7.6	420	1	US-08-504-718D-14

c 28	28.8	7.6	420	1	US-08-504-718D-15	Sequence 15, Appl
c 29	28.8	7.6	420	1	US-08-504-718D-16	Sequence 16, Appl
c 30	28.8	7.6	1205	4	US-09-017-734A-19	Sequence 19, Appl
c 31	28.4	7.5	2259	1	US-07-828-700-7	Sequence 7, Appl
c 32	28.4	7.5	2358	4	US-09-134-001C-1029	Sequence 1029, Ap
c 33	28.2	7.5	1068	1	US-08-427-640-7	Sequence 7, Appl
c 34	28.2	7.5	1068	1	US-08-981-803-9	Sequence 9, Appl
c 35	28.2	7.4	376	3	US-08-981-803-23	Sequence 23, Appl
c 36	28.2	7.4	376	4	US-08-983-440-9	Sequence 9, Appl
c 37	28.2	7.4	376	4	US-08-983-440-23	Sequence 23, Appl
c 38	28.2	7.4	699	4	US-09-221-017B-169	Sequence 169, App
c 39	28.2	7.4	11478	3	US-08-981-803-29	Sequence 29, Appl
c 40	28.2	7.4	11478	4	US-08-983-440-29	Sequence 29, Appl
c 41	27.6	7.3	2298	2	US-08-951-648-1	Sequence 1, Appl
c 42	27.6	7.3	2298	2	US-09-174-437-1	Sequence 1, Appl
c 43	27.6	7.3	3195	2	US-08-951-648-5	Sequence 5, Appl
c 44	27.6	7.3	3195	3	US-09-174-437-5	Sequence 5, Appl
c 45	27.6	7.3	4389	2	US-08-951-648-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-845-258-10
; Sequence 10, Application US/08845258
; Patent No. 6183976
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Iodes, Michael J.
; APPLICANT: Houghton, Raymond
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,258
; FILING DATE: 24-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MAKI, David J.
; REGISTRATION NUMBER: 31,392
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-845-258-10
Query Match 8.4%; Score 31.6; DB 4; Length 3701;
Best Local Similarity 51.4%; Pred. No. 1.6;
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 142 TTTTCGGTGTGTTTAAATGATGTTCTTTGGAATAAACCCTCCGCAAGTACTATG 201
DB 999 TATTCTACTTTTAAATATATTTTATTTATTCATATATATCTCTTTCATATTTTGTACTAT 1058

OY	202	GTTTATAGCCACATGCTTGACTGTGAAGAAGACCTTTGGCACCATGTAAGGAT	261
Dδ	1059	TTTATTAAATCATAATTATTATATATTTAATTTGATTAATGAATATATCAATPAT	1118
OY	262	GGTCATCGAGATGCCAATCCAT	283
Dδ	1119	GATGATATTACATCATATATGCAT	1140

RESULT 2

US-08-990-571-10
; Sequence 10, Application us/08990571
; Patent No. 6214871
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G. et al.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B. M
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,571
; FILING DATE: 11-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: MAKI, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4800
; TELEFAX: (206)682-6031
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-990-571-10

	Query Match	8.4%	Score 31.6;	DB 4;	Length 3701;
	Best Local Similarity	51.4%	Pred. No. 1.6;		
	Matches 73; Conservative	0;	Mismatches 69;	Indels 0;	Gaps 0;
QY	142 TTTTCCGTTGTTATCATGTCTCTTGGATAAAACCTCCCGCAAGTAGACTTG	201			
Dd	999 TATCTACTTTTATATATATTTTTATTCAATATATACCTTCATATTTTGTACTAT	1058			
QY	202 GTTTTATGCTTAACAATGCTTGTGAAGAGACCTTGGCACATVGAAGGAT	261			
Dd	1059 TTTTATATCATCATATATATATATATATATATATATTTGATTAATGAAATATATCATATAT	1118			
QY	262 GGTGATGAGATGCCAATCCAT	283			
Dd	1119 GATGATATACATGAAATATGCAT	1140			

RESULT 3

US-08-723-142A-10
Sequence 10, Application US/08723142A
Patent No. 6306396
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.

APPLICANT: Houghton, Raymond
APPLICANT: Sleath, Paul R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/723,142A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121,426
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206)682-6031
INFORMATION FOR SEQ. ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 3701 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-723-142A-10

[illegible]

RESULT 4

```

US-09-528-784A-10
: Sequence 10, Application US/09528784A
: Patent No. 6451315
:
: GENERAL INFORMATION:
:
: APPLICANT: Reed, Steven G.
:
: APPLICANT: Lodes, Michael J.
:
: APPLICANT: Houghton, Raymond L.
:
: APPLICANT: Sleath, Paul R.
:
: APPLICANT: McNeill, Patricia D.
:
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
:
: AND TREATMENT OF B. MICROIDI INFECTION
:
: FILE REFERENCE: 210121.426C4
:
: CURRENT APPLICATION NUMBER: US/09/528,784A
:
: CURRENT FILING DATE: 2000-03-17
:
: NUMBER OF SEQ. ID NOS: 90
:
: SOFTWARE: FastSeq for Windows Version 3.0
:
: SEQ ID NO 10
:
: LENGTH: 3701
:
: TYPE: DNA

```


Db 369 AGTTTCATGTTGGAGCGTTGCTGTAATCCACACCCCTCTGTGACGGGTTACAGGCTCC 310
QY 240 CTTTGCCACATTTGAAGGATGCTGATGAGATGCCAATCCATGAGAAATCAG 291
Db 309 AGTATGTTCTGATCTATGAGATCCTGCTGACACTCCCTCCCTGGAAGCTG 258

RESULT 8

US-09-006-353A-1/c
Sequence 1, Application US/09006353A
Patent No. 6261801
GENERAL INFORMATION:
APPLICANT: WEI, YING-FEI
APPLICANT: YU, GUO-LIANG
APPLICANT: GENTZ, REINER
APPLICANT: RUBEN, STEVEN
TITLE OF INVENTION: TUMOR NECROSIS FACTOR RECEPTOR 5
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUMAN GENOME SCIENCES, INC.
STREET: 9410 KEY WEST AVENUE
CITY: ROCKVILLE
STATE: MD
COUNTRY: US
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,353A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BROOKES, ANDERS A
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF341
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1392 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 183..959
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 183..260
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 261..959
US-09-006-353A-1

Query Match Best Local Similarity 7.8%; Score 29.6; DB 4; Length 1392;
Matches 83; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 120 CACGGTTGTAAGCCCCCTGTTCCGTTGTTAATCTAATGTTCTTTGGAATAAAAA 179
Db 491 CATGGGACGAGAACTTTATGTTTGATCTGATTTACAACCTGATACATGGAAGCAAGA 432
QY 180 CCTCCCTGCAAGTACTAGTGGTTTATGCTCAACATGCTTGACTGTGAAAGAGAC 239
Db 431 AGTTTCATTTGTTGAAGCGTTGGTGAATCCACACCCCTCTGTGACGGGTTACAGGCTCC 372
QY 240 CTTTGCCACATTTGAAGGATGCTGATGAGATGCCAATCCATGAGAAATCAG 291

Db 371 AGTATGTTCTGATCTATGAGATCCTGCTGACACTCTCCCTTGAAGCTG 320

RESULT 9

US-09-573-986-1/c
Sequence 1, Application US/09573986
Patent No. 6455040
GENERAL INFORMATION:
APPLICANT: WEI, YING-FEI
APPLICANT: NI, JIAN
APPLICANT: GENTZ, REINER
APPLICANT: RUBEN, STEVEN
TITLE OF INVENTION: Tumor Necrosis Factor Receptor 5
FILE REFERENCE: 1488.1280004
CURRENT APPLICATION NUMBER: US/09/573,986
CURRENT FILING DATE: 2000-05-18
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 1392
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (183)..(959)
US-09-573-986-1

Query Match Best Local Similarity 7.8%; Score 29.6; DB 4; Length 1392;
Matches 83; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 120 CACGGTTGTAAGCCCCCTGTTCCGTTGTTAATCTAATGTTCTTTGGAATAAAAA 179
Db 491 CATGGGACGAGAACTTTATGTTTGATCTGATTTACAACCTGATACATGGAAGCAAGA 432
QY 180 CCTCCCTGCAAGTACTAGTGGTTTATGCTCAACATGCTTGACTGTGAAAGAGAC 239
Db 431 AGTTTCATTTGTTGAAGCGTTGGTGAATCCACACCCCTCTGTGACGGGTTACAGGCTCC 372
QY 240 CTTTGCCACATTTGAAGGATGCTGATGAGATGCCAATCCATGAGAAATCAG 291
Db 371 AGTATGTTCTGATCTATGAGATCCTGCTGACACTCTCCCTTGAAGCTG 320

RESULT 10

US-08-687-289A-2
Sequence 2, Application US/08687289A
Patent No. 5981195

GENERAL INFORMATION:
APPLICANT: FULLER, Forrest H.
APPLICANT: KRAPCHO, Karen J.
TITLE OF INVENTION: CHIMERIC RECEPTORS AND METHODS FOR IDENTIFYING COMPOUNDS ACTIVE AT METABOTROPIC GLUTAMATE RECEPTORS AND THE USE OF SUCH COMPOUNDS IN THE TREATMENT OF NEUROLOGICAL DISORDERS
TITLE OF INVENTION: AND DISEASES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSeq for Windows 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/687, 289A
FILING DATE: July 25, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/001,526
FILING DATE: July 26, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 220/004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4000 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-687-289A-2

Query Match

7.8%; Score 29.6; DB 2; Length 4000;

Best Local Similarity 50.7%; Pred. No. 7.5; Mismatches 69; Indels 0; Gaps 0;

Matches 71; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 35 GCACGGCTCTTAGACATATGAAACACCCAGCGGAGTCTCACAAGCTTGA 94

Db 1738 GAGAGTCTCTGATGGGAGTATAGAGAGACAGATGCGCTGTACAGAGCCG 1797

QY 95 ATGTGTCTTGGAGCTGAAGAGTACAGCGTTGTTAAGCCCTGTTTCCGTTGTT 154

Db 1798 AGATGACTTCTGCTCAATGAGAACCACTCGCGAGCCCATTCGTCCGTATCT 1857

QY 155 AATCAATGCTCTTGAAT 174

Db 1858 TGAGTGAGTACATAGAT 1877

RESULT 11

US-09-209-668-18

Sequence 18, Application US/09209668A

Patent No. 6114517

GENERAL INFORMATION:

APPLICANT: Monla, Brett P.

APPLICANT: Xu, Xiaojing S.

TITLE OF INVENTION: METHODS OF MODULATING TUMOR NECROSIS FACTOR

FILE REFERENCE: ISPH-0336

CURRENT APPLICATION NUMBER: US/09/209,668A

CURRENT FILING DATE: 1998-12-10

NUMBER OF SEQ ID NOS: 25

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 18

LENGTH: 3834

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (117)..(1949)

PUBLICATION INFORMATION:

DATABASE ACCESSION NUMBER: M24736/Genbank

DATABASE ENTRY DATE: 1994-11-07

US-09-209-668-18

Query Match

7.8%; Score 29.4; DB 3; Length 3834;

Best Local Similarity 49.1%; Pred. No. 8.6;

Matches 78; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 114 AGGAGCAGCGTTGTAAGCCCGTCTTCCGTTTAACTAAATCTTTGGAA 173

Db 3192 AAGATGCTAATAGTATTCCTATTGTTTCTCTGTATGAGGTCCTGGAAGA 3251

QY 174 TAAACCTCCGCAAGTACTGTTTATGCTCAACAGCTTGCCTTGGA 233
Db 3252 GAGGAATGCTGTGTGAGCAAGCATTTATTTATTAAGCATTTAAACATTCGA 3311

QY 234 AGAGACCTTGGCAGCATTTGAAGGATGCTGAGGA 272
Db 3312 AGGAATCTCAAGTTTCAGTTGATCACTGGCAATGAAA 3350

RESULT 12

US-08-365-470-1

Sequence 1, Application US/08365470

Patent No. 5632991

GENERAL INFORMATION:

APPLICANT: Gimbione, Jr., Michael A.

TITLE OF INVENTION: Antibodies Specific For E-selectin And the Uses

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSER: STERNE, KESSLER, GOLDSTEIN & FOX

STREET: 1100 New York Ave., NW

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/365,470

FILING DATE: herewith

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/102,510

FILING DATE: 05-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/850,802

FILING DATE: 13-MAR-1992

ATTORNEY/AGENT INFORMATION:

NAME: Markowicz, Karen R.

REGISTRATION NUMBER: 36,351

REFERENCE/DOCKET NUMBER: 0627.1350003

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-2540

TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3854 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-365-470-1

Query Match

7.8%; Score 29.4; DB 1; Length 3854;

Best Local Similarity 49.1%; Pred. No. 8.6;

Matches 78; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 114 AGATGACGCTTGTATAGCCCGTCTTCCGTTTAACTAAATGCTTTGGA 173

Db 3192 AAGATGCTAATAGTATTCCTATTGTTTCTCTGTATGAGGTCCTGGAAGA 3251

QY 174 TAAACCTCCGCAAGTACTGTTTATGCTCAACAGCTTGCCTTGGA 233

Db 3252 GAGGAATGCTGTGTGAGCAAGCATTTATTTATTAAGCATTTAAACATTCGA 3311

QY 234 AGAGACCTTGGCAGCATTTGAAGGATGCTGAGGA 272

Db 3312 AGGAATCTCAAGTTTCAGTTGATCACTGGCAATGAAA 3350

QY 234 AGAGACCTTTGGCACACATTTGAAGGAGTGTGATGGAGA 272
 Db 3336 AGGAATCTCCAGTTTTCAGTTGATCATCTGCAATGAAA 3374

RESULT 15
 US-08-482-073-1
 ; Sequence 1, Application US/08482073
 ; Patent No. 6307025
 ; GENERAL INFORMATION:
 ; APPLICANT: Hession, Catherine A.
 ; APPLICANT: Lobb, Roy R.
 ; APPLICANT: Goetz, Susan E.
 ; APPLICANT: Osborn, Laurelee
 ; APPLICANT: Benjamin, Christopher D.
 ; APPLICANT: Rose, Margaret D.
 ; TITLE OF INVENTION: ENDOTHELIAL CELL-LEUKOCYTE ADHESION
 ; TITLE OF INVENTION: MOLECULES (ELAMS) AND MOLECULES INVOLVED IN LEUKOCYTE
 ; NUMBER OF INVENTIONS: ADHESION (MILAS)
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Neave
 ; STREET: 1251 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States of America
 ; ZIP: 10020
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/482.073
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/486.336
 ; FILING DATE:
 ; APPLICATION NUMBER: US 07/608298
 ; FILING DATE: 31-OCT-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/US 90/02357
 ; FILING DATE: 27-APR-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/452675
 ; FILING DATE: 18-DEC-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/359516
 ; FILING DATE: 01-JUN-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/354151
 ; FILING DATE: 28-APR-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Haley Jr., James F.
 ; REGISTRATION NUMBER: 27,794
 ; REFERENCE/DOCKET NUMBER: B124CIP4
 ; TELEPHONE: (212) 596-9000
 ; TELEFAX: (212) 596-9090
 ; TELEX: 14-8367
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3863 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; US-08-482-073-1

Query Match 7.88; Score 29.4; DB 4; Length 3863;
 Best Local Similarity 49.18; Pred. No. 8.6;
 Matches 78; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 114 AGGATGACAGGCTTTTAAGCCCTGTTTCCTTTTCCGTTTAACTAATGTCTTTGAA 173
 Db 3215 AGATGTCTAATAGTTATTCCTATTTGTTTCTGTATGTTAAGGTCCTCTGGAAGA 3274
 QY 174 TAAAAACCTCCCGCCAGAGTACTGTTTATGCTCAACATGCTTGCATGTTGAA 233
 Db 3275 GAGGATGCGCTGTGTAGAGCAGCATTTATTTATTAAGCAGATTTAACAATTCCAA 3334
 QY 234 AGAGACCTTTGGCACACATTTGAAGGAGTGTGATGGAGA 272
 Db 3335 AGGAATCTCCAGTTTTCAGTTGATCATCTGCAATGAAA 3373

Search completed: January 31, 2003, 01:14:56
 Job time : 88 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2003, 01:17:19 ; Search time 2405 Seconds

(without alignments)
2545.486 Million cell updates/sec

Title: US-09-803-719-222

Perfect score: 378
Sequence: 1 tacgcgtgcttaagacgact.....taccattgagaacatcgcg 378

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```
1: em_estba:*
2: em_esthum:*
3: em_estfn:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	167.8	44.4	451	14	BM987874 UI-H-COO-
2	118.8	31.4	439	9	AI459918
3	108.8	28.8	386	9	AA743908
4	79.4	21.0	806	12	BG197212
5	62.6	16.6	349	17	AQ035618
6	41.4	11.0	237	17	AQ035738

7	39.8	10.5	554	14	BQ655384
8	39.2	10.4	439	17	BH853295
9	38.8	10.3	1237	13	BI454874
10	38	10.1	585	14	BQ392762
11	38	10.1	641	9	AL628007
12	37.4	9.9	592	17	AQ922388
13	37	9.8	699	17	A2108808
14	37	9.8	829	17	A2191123
15	36.8	9.7	473	9	AT449346
16	36.4	9.6	550	17	AQ775948
17	36.4	9.6	798	9	AI350894
18	36	9.5	254	9	AI600177
19	35.2	9.3	601	10	AW704405
20	35	9.3	807	17	CNS00AMP
21	34.8	9.2	398	9	AT181854
22	34.8	9.2	554	17	A2720218
23	34.8	9.2	695	14	BM973205
24	34.6	9.2	246	10	BB566524
25	34.6	9.2	532	17	A2182814
26	34.6	9.2	935	13	BI861604
27	34.4	9.1	416	10	AV413421
28	34.4	9.1	728	17	A2400532
29	34.4	9.1	819	17	CNS0440P
30	34.2	9.0	562	10	AW231608
31	34	9.0	649	10	BB481722
32	33.8	8.9	571	9	AT828325
33	33.8	8.9	584	17	AQ419832
34	33.8	8.9	643	10	AW190856
35	33.8	8.9	654	17	A2083843
36	33.6	8.9	345	10	AW444676
37	33.6	8.9	491	14	BQ024523
38	33.6	8.9	521	17	A2182824
39	33.4	8.8	247	10	BE067501
40	33.4	8.8	286	9	AI651773
41	33.4	8.8	337	9	AA711119
42	33.4	8.8	499	10	AW635726
43	33.4	8.8	685	13	BQ068145
44	33.4	8.8	899	17	A2534719
45	33.2	8.8	488	12	BG136484

ALIGNMENTS

RESULT 1
BM987874/c 451 bp mRNA linear EST 17-JUN-2002
UI-H-COO-asz-a-11-0-UI.s1 NCI_CGAP_Sub9 Homo sapiens cDNA clone
IMAGE:5860149 3', mRNA sequence.
BM987874
BM987874.1 GI:19707263

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgap@remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
1. .451

FEATURES
source

a Modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 93 a 77 c 73 g 143 t

Query Match 28.8%; Score 108.8; DB 9; Length 386;
Best Local Similarity 98.2%; Pred. No. 6.3e-21;
Matches 110; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 266 ATGGAGATCCCAATCCATGAGATCAGTGGCCAGCTATGTTGGTACATACAGCACT 325
|||||
386 ATGGAGATCCCAATCCATGAGATCAGTGGCCAGCTATGTTGGTACATACAGCACT 327

Qy 326 CTTCTGGCAAAAGATTCTCCCGGAGAGAGTACATTTGGAGAACATGCG 377
|||||
Db 326 CTTCTGGCAAAAGATTCTCCCGGAGAGAGTACATTTGGAGAACATGCG 275

RESULT 4
Bg197212/c 806 bp mRNA linear EST 21-APR-2001
LOCUS Bg197212
DEFINITION Bg197212 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION Bg197212
VERSION Bg197212.1 GI:13718899
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 806)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramchandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Boozar,S., Mays,R., Smith,
E., Veloso,N., Kilka,A., Hess,J., Cothran,K., Lo,K., Offenbacher,
J., Danzig,J. and Ducar,M.

TITLE Creation of genome-wide protein expression libraries using random
activation of gene expression
JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE 21227151
COMMENT Contact: Scott J. Cain

Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: ecain@atersys.com

FEATURES
source High quality sequence stop: 417.
Location/Qualifiers

1..806
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."
BASE COUNT 252 a 187 c 155 g 205 t 7 others

Query Match 21.0%; Score 79.4; DB 12; Length 806;
Best Local Similarity 98.8%; Pred. No. 2.6e-12;
Matches 80; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 262 GGTGATGAGATGCCAATCCATGAGATCAGTGGCCAGCTATGTTGGTACATACAG 321
|||||

Db 423 GGTGATGAGATGCCAATCCATGAGATCAGTGGCCAGCTATGTTGGTACATACAG 364

Qy 322 AAGCTTTCTGGCAAAAGATTTC 342
|||||
Db 363 AAGCTTTCTGGCAAAAGATTTC 343

RESULT 5
A0035618 349 bp DNA linear GSS 11-JUL-1998
LOCUS A0035618
DEFINITION CIT-HSP-2319N3.TF CIT-HSP Homo sapiens genomic clone 2319N3, DNA sequence.
ACCESSION A0035618
VERSION A0035618.1 GI:3301715
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 349)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.

TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other GSSs: CIT-HSP-2319N3.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
source Location/Qualifiers
1..349
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="2319N3"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; site_1: HindIII; site_2: HindIII"

BASE COUNT 85 a 69 c 84 g 111 t

Query Match 16.6%; Score 62.6; DB 17; Length 349;
Best Local Similarity 94.2%; Pred. No. 1.7e-07;
Matches 65; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 86 CAAGCTTGATGTGTCTTCGAGCTGAAGATGACACGGTTGTAACCCCTGTCTTTT 145
|||||
Db 31 CAAGCTTGATGTGTCTTCGAGCTGAAGATGACACGGTTGTAACCCCTGTCTTTT 90

Qy 146 CCGTTGTTT 154
|||||
Db 91 CCGTAAGTT 99

RESULT 6
A0035738 237 bp DNA linear GSS 11-JUL-1998
LOCUS A0035738
DEFINITION CIT-HSP-2321C21.TF CIT-HSP Homo sapiens genomic clone 2321C21, DNA sequence.
ACCESSION A0035738
VERSION A0035738.1 GI:3301835
KEYWORDS GSS.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 237)
AUTHORS Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M., and Venter, J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2321C21.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: madams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES
SOURCE
1. .237
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2321C21"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelBAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT
ORIGIN
59 a 47 c 47 g 84 t

Query Match
Best Local Similarity 97.7%; Score 41.4; DB 17; Length 237;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 335 AAGATTCCTCCCGAAGAGGATGACATGAGAACCATGC 377
|||||
Db 237 AAGATTCCTCCAGGAAGAGGATGACATGAGAACCATGC 195
|||||

RESULT 7
LOCUS B0653384 554 bp mRNA linear EST 15-JUL-2002
DEFINITION NXRV093.F11.F NXRV (Nsf Xylem Root wood Vertical) Pinus taeda cDNA
ACCESSION B0653384
VERSION B0653384.1 GI:21787710
KEYWORDS EST.
SOURCE loblobly pine.
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 554)
AUTHORS Sederoff, R.
TITLE Molecular Basis of Wood Formation in the Pine Megagenome
JOURNAL Unpublished (2000)
COMMENT Contact: Johnson, Arthur
North Carolina State University
Tel: 919 515 7800
Fax: 919 515 7801
Email: ajohnson@unity.ncsu.edu
Seq primer: T3.

FEATURES
SOURCE
1. .554
/organism="Pinus taeda"
/strain="Coastal plain loblobly pine from North Carolina"
/db_xref="taxon:3352"

/clone="NXRV093.F11"
/clone_lib="NXRV (Nsf Xylem Root wood Vertical)"
/cissue_type="Xylem"
/cell_type="Root (primary)"
/dev_stage="Transitional"
/lab_host="XL1-Blue"
/note="Vector: pBluescript SK-; Site_1: Eco RI; Site_2: XhoI; The library is from primary xylem scraped from the roots of a twelve year old tree in the transitional phase from juvenile wood to mature wood production. NOTE: The sequences contain a 'cDNA adapter' between the EcoRI site and the start of the EST. The adapter sequence is 'ATTTCGCACGAG'."

BASE COUNT
ORIGIN
153 a 88 c 131 g 161 t 21 others

Query Match
Best Local Similarity 10.5%; Score 39.8; DB 14; Length 554;
Matches 83; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 168 TTGGAATTAACCTCCCTCCAGATGACTGTTTATGCTCAACATGCTTGACTG 227
|||||
Db 287 TTGCAAGAAAACATTTACTGCAATTTGCTGTTTATGCTGATTTGTTGAATG 346
|||||

QY 228 TTGAAGAGAGACCTTTGGCAGACATTTGAAGGATGCTGATGAGATGCCAATTCATGAA 287
|||||

Db 347 CAGATATATGTTCTTCCAGAAATACATACATGAGGATGAGGACGGGTAAAGCAATTAATGAA 406
|||||

QY 288 TCAGGTGGCGACGATGATGTTGTTGATGCTATGACAG 322
|||||

Db 407 TGAATCTGGAGCGTTTGATGATGCTTCAATTTGATAGA 441
|||||

RESULT 8
LOCUS BH853295 439 bp DNA linear GSS 13-JUN-2002
DEFINITION SALK_076311.54.00.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_076311.54.00.x, DNA sequence.
ACCESSION BH853295
VERSION BH853295.1 GI:21424166
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 439)
AUTHORS Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Preditsis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished (2001)
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 X1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 3' end of At2g41810.
Class: TDNA tagged.

FEATURES
SOURCE
1. .439
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_076311.54.00.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines

each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/telna_protocols.html.

BASE COUNT

135 a 68 c 68 g 146 t 2 others

Query Match 10.4% Score 39.2; DB 17; Length 439;
Best Local Similarity 53.2%; Pred. No. 1.1;
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 106 GGAGCTGAAGATGCAAGGCTGTTAAAGCCCTGTTCTTTCCGTTGTTAATCTAATGT 165
DB 123 GGAGACCATATATGACTGTGCAACCTCCCTTAACATATCTCTCGTACACACATGCC 182
QY 166 CTTTGAATATAAAACCTCCCTGCAAGTACTGTTTAAATGCTCAACATCTTTGAC 225
DB 183 CATTTAATATAAATCTTATTTATGACACAAATTTCTTTTAAAGTTTCTTTGTA 242
QY 226 TGTGAAAGACAGCTTTGGACACACATTTGAAGGAT 261
DB 243 AGTTAAAGCTAAAGCTTTTATTTAAACAAGACAAAT 278

RESULT 9

LOCUS B1454874 1237 bp mRNA linear EST 21-AUG-2001
DEFINITION 603173414F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5252883 5',
mRNA sequence.

ACCESSION B1454874
VERSION B1454874
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE NIH-MGC http://imgc.ncl.nih.gov/
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 1237)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://imgc.ncl.nih.gov>
Plate: L1AM1638 row: 1 column: 04
High quality sequence stop: 102.

FEATURES

source 1. 1237
Location/Qualifiers

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="5252883"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1; Salt: Site: 2; Note: Cloned unidirectionally. Primer: Oligo dt. library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"

BASE COUNT 431 a 226 c 336 g 243 t 1 others

Query Match 10.3% Score 38.8; DB 13; Length 1237;
Best Local Similarity 50.3%; Pred. No. 1.8;
Matches 94; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 65 CACCAAGCGGAGTCTCTCACAAGCTTGAATGTGTCTTGGAGCTGAAGATGACACG 124
DB 480 CCCCCCTCCCTCTCTCTTTGTTGTGTTGTTGTTGTTCTTCTGTATC 421
QY 125 TTGTTAAGCCCTGTTCTTTCCGTTGTTAATGTTCTTTGAAATAAACCTCC 184
DB 420 TTCTCTATCTGTTATTTGTTTCTTTTGTATCTGTGTTTCTTTTATATAGACA 361
QY 185 CTGCAAGTAGTACTGTTTAACTCAACATCTTTGACTTTGAAAGAGACTTTG 244
DB 360 CAGCAATATATTTTATTTATGAGACACACACTGTTTTCACATTTTACACCATTTGG 301
QY 245 GCACACA 251
DB 300 GCGCACA 294

RESULT 10

LOCUS B0392762 585 bp mRNA linear EST 22-MAY-2002
DEFINITION B0392762
IMAGE:5384863 5', mRNA sequence.

ACCESSION B0392762
VERSION B0392762
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis

REFERENCE NIH-XGC http://imgc.ncl.nih.gov/image/hm1/xenopuslib_info.shtml.
AUTHORS National Institute of Child Health and Human Development, National
Cancer Institute, Xenopus Gene Collection
TITLE Unpublished (2002)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
cDNA Library Preparation:
DNA Sequencing by: The I.M.A.G.E. Consortium/LNL
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://imgc.ncl.nih.gov>
Plate: L1AM1980 row: G column: 8
Seq primer: M13RPI reverse primer (AB1).

FEATURES

source 1. 585
Location/Qualifiers

/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone_image="5384863"
/clone_lib="NICHG XGC Emb5"
/tissue_type="gastrula"
/dev_stage="embryo, stages 10-13"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pCMV-SPORT6, codb; Site: 1; Note: Site: 2;
ECORV; Cloned unidirectionally. Primer: Oligo dt. Average
insert size 2.0 kb. Constructed by Invitrogen. Note: This
is a Xenopus Gene Collection (XGC) library."

BASE COUNT 141 a 109 c 163 g 172 t

Query Match 10.1% Score 38; DB 14; Length 585;
Best Local Similarity 52.5%; Pred. No. 2.6;
Matches 83; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 144 TTCCGTTGTTTAACTAATGTTCTTTGAAATAAACCTCCGCAAGTAGTACTTGGT 203
DB 405 TTGCCATGAGAGACTCTGATTTCTTGGAATGAAAGGCTTACTAGAGCTGTGACA 464
QY 204 TTATCTCAACATGCTTGAAGGAGCTTTGGACACATTTGAAGGATGG 263

Db 465 CCTCTCTTTGAGAGTTTGCCTTAATTAAGAGACTTGTGGACACCAAGCAAGACGAA 524

Qy 264 TGATGAGATGCCAATTCATGCATGAGTGGCGACG 301

Db 525 TCATAGTAATGCCAGTGGCGCATTTACTGACGGAAC 562

RESULT 11

AL628007 641 bp mRNA linear EST 02-NOV-2001

DEFINITION AL628007 XGC-gastrula Silurana tropicalis cDNA clone Tgas022e21 5', mRNA sequence.

ACCESSION AL628007

VERSION AL628007.1 GI:16597490

KEYWORDS western clawed frog.

SOURCE Silurana tropicalis

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;

REFERENCE Huckle, E., Taylor, R., Ashurst, J. L., Zorn, A. M. and Rogers, J. 1 (bases 1 to 641)

AUTHORS Sanger Xenopus tropicalis EST project 2001 (10_2001)

TITLE Unpublished (2001)

JOURNAL Contact: Huckle E

COMMENT Sanger Centre

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: tropesanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS_SEQUENCE_ID: Tgas022e21.sp6

Sequencing primer: Sp6

This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.

FEATURES

source

Location/Qualifiers

1. 641

/organism="Silurana tropicalis"

/db_xref="taxon:8364"

/clone="Tgas022e21"

/dev_stage="XGC-gastrula"

/lab_host="Escherichia coli XL1-blue"

/note="Vector: PCS107; Site_1: EcoRI; Site_2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into PCS107 with EcoRI at the 5' end and NotI at the 3' end."

BASE COUNT 154 a 118 c 189 g 180 t

ORIGIN

Query Match 10.1% Score 38; DB 9; Length 641;

Best Local Similarity 52.5% Pred. No. 2.7; Mismatches 83; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 144 TTCCTGTTTAATCTAATGTTCTTTGGAATAAAACCTCCGCCAATGACTGTGT 203

Db 449 TTGCATGAGAGCTCTGATTTCTTGGAGATTTGAAAGGTTTACTAGTGGCTGACA 508

Qy 204 TTATGCTCAACATGCTTGTGACTGTGAAAAGACCTTTGGCACACATTTGAAGGATG 263

Db 509 CCTCTGCTTTGAGAGTGTGCTTGAATTAAGAGACTTGTGAGACCAAGCAAGACGAA 568

Qy 264 TGATGAGATGCCAATTCATGCATGAGTGGCGACG 301

Db 569 TCATAGTAATGCCAGTGGCGCATTTACTGACGGAAC 606

RESULT 12

LOCUS A0922388 592 bp DNA linear GSS 21-DEC-1999

DEFINITION RPCI-23-276B5.TJ RPCI-23 Mus musculus genomic clone RPCI-23-276B5, DNA sequence.

ACCESSION A0922388

VERSION A0922388.1 GI:6611391

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 592)

AUTHORS Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akintet, B., Levins, M., McGann, S., Tesgaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C. M.

TITLE Mouse BAC End Sequences from Library RPCI-23

JOURNAL Unpublished (1999)

COMMENT Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhaoc@igf.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.html>) or from Resea.ch Genetics (http://www.tigr.org/tdb/Bac_ends/mouse/Bac_end_intro.html)

plate: 276 row: B column: 5

Seq primer: Sp6

Class: BAC ends.

FEATURES

source

Location/Qualifiers

1. 592

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-276B5"

/clone_1ib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylease. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 161 a 115 c 130 g 186 t

ORIGIN

Query Match 9.9% Score 37.4; DB 17; Length 592;

Best Local Similarity 50.3% Pred. No. 3.9; Mismatches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Qy 109 GCTGAAGATGCAACGCTTGAAGCCCTGTTCTTTCCGTTGTTTAATCTAATGTTCTT 168

Db 398 GTTGAATTTGCTCTTGACATGCTGTTCTTTTCCACAGTTTATCCACTTGTG 457

Qy 169 TGAATTAATAAACCTCCCTGCCCAAGTAGTACTGTTTATGCTCAACATGCTTGAAGT 228

Db 458 TTAATTAACATGCGCCAGCTTAAATTTGGAGCGCTTTGGGAGAAATTAATTTGTATG 517

Qy 229 TGAAGAAGACCTTTGGACACATTTGAAGGATGTTGATGAGATGCCAATCCATGGAAT 288

Db 518 GCCCGAGGCGGGGAGACAGTAAGTGGGTCTTGGAGATGGTAGTATTTCTATGTTCT 577

Qy 289 CAG 291

Db 578 CAG 580

RESULT 13

LOCUS A2108808 699 bp DNA linear GSS 09-MAY-2000

DEFINITION RPCI-23-25N18.TV RPCI-23 Mus musculus genomic clone RPCI-23-25N18, DNA sequence.

ACCESSION A2108808

VERSION	AZ108808.1			GI:7761866
KEYWORDS	GSS.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 699)			
	Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akhret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.			
TITLE	Mouse BAC End Sequences from Library RPCI-23			
JOURNAL	Unpublished (1999)			
COMMENT	Other_GSSs: RPCI-23-25N18.TJ Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@eijong.med.buffalo.edu). Clones may be purchased from BACpac Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Reser ch Genetics (Info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html Plate: 25 row: N column: 18 Seq primer: T7 Class: BAC ends			
FEATURES	Location/Qualifiers			
source	1..699			
	/organism="Mus musculus"			
	/strain="C57BL/6J"			
	/db_xref="taxon:10090"			
	/clone="RPCI-23-25N18"			
	/clone_1bp="RPCI-23"			
	/sex="Female"			
	/lab_host="DH10B"			
	/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."			
BASE COUNT	241 a 114 c 129 g 215 t			
ORIGIN				
Query Match	9.8%; Score 37; DB 17; Length 699;			
Best Local Similarity	56.0%; Pident. NO. 5.3;			
Matches	70; Conservative 0; Mismatches 55; Indels 0; Gaps 0;			
OY	121	ACGGTGTAAAGCCCTCTCTTTCCCTGTAAATGCTTGGAAATAAAC	180	
Db	538	ACTGCTGTTCCTCCGCTCGAGCTCTGATTAAGTAAATTATTAGATGTA	479	
OY	181	CTCCCTGCCAAGTAGTACTGTGGTTTATGCTCAACATGCTTGA	240	
Db	478	CTTAGTCACATTAAGATATTATCTCTGCACAAATTTCTTATT	419	
OY	241	TTTGG 245		
Db	418	TTTGG 414		
RESULT 14				
AZ191123/c	829 bp DNA linear GSS 30-AUG-2000			
LOCUS	SP_1018.B2.D06.T7A Strongyloides purpuratus, purple sea urchin			
DEFINITION	clone p1018-1018 Col-12 Row-H. DNA sequence.			
ACCESSION	AZ191123			
VERSION	AZ191123.1 GI:8374302			

REFERENCE	AUTHORS	TITLE
KEYWORDS	GSS.	Strongylocentrotus purpuratus.
SOURCE	Strongylocentrotus purpuratus	
ORGANISM	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidae; Euechinozoa; Echinacea; Echinoidae; Strongylocentrotidae; Strongylocentrotus.	
REFERENCE	1 (bases 1 to 829)	
AUTHORS	Cameron, R.A., Mahalinas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, G.A., Eitelsohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L.	
	A sea urchin genome project: Sequence scan, virtual map, and additional resources	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)	
MEDLINE	20402566	
COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 1018 row: H column: 12 Seq primer: T7 Class: BAC ends High quality sequence stop: 829.	
FEATURES	Location/Qualifiers	
source	1..829	
	/organism="Strongylocentrotus purpuratus"	
	/db_xref="taxon:7668"	
	/clone_plate="1018 Col-12 Row-H"	
	/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"	
	/note="Organ: sperm; Vector: BAC#3.6; BAC clones in E-Coli DH10B"	
BASE COUNT	258 a 190 c 151 g 229 t 1 others	
ORIGIN		
Query Match	9.8%; Score 37; DB 17; Length 829;	
Best Local Similarity	54.5%; Pred. No. 5.5;	
Matches	73; Conservative 0; Mismatches 61; Indels 0; Gaps 0;	
QY	151 GTTTATGTAAGTCTTTGGAAATAAAACCTCCCTGCCAGTACTGTTTATGC 210	
DB	641 GTTATGTAAGTCTTTGGTCTGAAAGAACCTTGAGGTATGCTTGAATATCC 582	
QY	211 TCACATGCTTGGACCTTGGAAAGAGACCTTGGACACATTTGAAGGATGATGGA 270	
DB	581 GCCACGCTATTCATATGTTGAACACTTCTCGGAACTTACGAGGAGGAGGGA 522	
QY	271 GATGCCAATCCATG 284	
DB	521 GCTAGCAACGCTG 508	
RESULT 15		
LOCUS	A1449346	
DEFINITION	mt07d11.x1 Soares mouse 3NBMS Mus musculus CDNA clone IMAGE:620373	
ACCESSION	A1449346	
VERSION	A1449346	
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 473)	
	Marras, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, T., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.	

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2003, 22:59:19 ; Search time 263 Seconds

(without alignments)
3236,712 Million cell updates/sec

Title: US-09-803-719-222

Sequence: 1 tacggctgcttaagacgact.....taccattggagaacatgcg 378

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

_N_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB	ID	Description
1	378	100.0	378	22	AA537164
2	299.6	79.3	22	22	AA57218
3	279.6	74.0	22	22	AA57237
4	275	72.8	22	22	AA57242
5	33.8	8.9	24	24	AB133878
6	33.6	8.9	22	22	ABA08523
7	33.4	8.8	21	21	AAC29807
8	33.4	8.8	20	20	AA224099
9	33.4	8.8	20	20	AA224100

C	10	33.4	8.8	25509	21	AAA81467	N. meningitidis pa
C	11	33.4	8.8	34980	21	AA521607	Neisseria meningit
C	12	33	8.7	303	22	AAK57669	Human immune/haema
C	13	32.8	8.7	240	20	AA224101	'Rat NF-1 DNA fragm
C	14	32.8	8.7	240	20	AA224102	Rat NF-1 DNA fragm
C	15	32.8	8.7	240	20	AA224103	Rat NF-1 DNA fragm
C	16	32.8	8.7	240	20	AA224104	Rat NF-1 DNA fragm
C	17	32.8	8.7	240	20	AA224105	Rat NF-1 DNA fragm
C	18	32.4	8.6	413	22	AAK53947	Murine translation
C	19	32.2	8.5	1445	16	AA090508	Tomato S-adenosyl-
C	20	32	8.5	410	21	AAK26229	Human secreted pro
C	21	32	8.5	490	24	AB060002	Human colon cancer
C	22	32	8.5	1604	22	AAAF84752	Nucleotide sequenc
C	23	31.8	8.4	240	20	AA224098	Rat NF-1 DNA fragm
C	24	31.8	8.4	314	22	AAK70860	Human immune/haema
C	25	31.8	8.4	606	23	ABV06997	Human prostate exp
C	26	31.8	8.4	3738	23	ABL12954	Drosophila melanog
C	27	31.8	8.4	13317	22	AA546681	Tumour suppressor
C	28	31.8	8.4	13317	24	AA561366	Human gene regulat
C	29	31.6	8.4	1003	23	AA579681	DNA encoding novel
C	30	31.6	8.4	2966	24	ABL89380	Babesia microti an
C	31	31.6	8.4	3011	24	ABL89389	Babesia microti an
C	32	31.6	8.4	3701	19	AAV22740	Babesia microti an
C	33	31.6	8.4	3701	20	AAV90004	Babesia microti an
C	34	31.6	8.4	3701	21	AAK5087	Babesia microti an
C	35	31.6	8.4	3701	24	AB189331	B. microti BMN1-10
C	36	31.6	8.4	580073	18	AA558840	Babesia microti an
C	37	31.4	8.3	568	22	ABA61524	Mycoplasma genital
C	38	31.4	8.3	568	22	AAK09824	Human foetal liver
C	39	31.4	8.3	568	22	AAK35718	Human bone marrow
C	40	31.4	8.3	568	22	AA141433	Probe 10119 used
C	41	31.4	8.3	568	24	AB509989	Human genome-deriv
C	42	31.4	8.3	13319	23	AA587926	DNA encoding novel
C	43	31.2	8.3	40324	24	AB067149	Human anglogenesis
C	44	31	8.2	362	21	AA67558	Eucalyptus grandis
C	45	30.8	8.1	470	23	ABV09451	Human prostate exp

ALIGNMENTS

RESULT 1	AA537164	standard; cDNA; 378 BP.
ID	AA537164	
AC	AA537164	
XX		
XX		
DT	17-DEC-2001	(first entry)
DE	Novel human diagnostic and therapeutic gene #222.	
XX		
KW	Human; cancer; breast; lung; colon; prostate; cytosolic; diagnostic; ss.	
XX		
OS	Homo sapiens.	
PN	WO200166753-A2.	
PD	13-SEP-2001.	
PF	09-MAR-2001; 2001WO-US07787.	
PR	09-MAR-2000; 2000US-0188609.	
PA	(CHIR) CHIRON CORP.	
PA	(HYSE-) HYSEQ INC.	
PI	Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;	
PI	Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;	
PI	Drimnac R, Ctkventjakov R, Dickson M, Drimnac S, Labat I;	
PI	Leskowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;	
DR	WPI, 2001-530177/58.	
XX		

PT New polynucleotides and polypeptides, useful for diagnosis and
treatment of breast, lung and colon cancer.

PS Claim 1; Page 652; 1193pp; English.

XX The invention relates to new polynucleotides and polypeptides, useful for
PS diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.

XX Sequence 378 BP; 100 A; 76 C; 101 G; 101 T; 0 other;

SO Query Match 100.0%; Score 378; DB 22; Length 378;
Best Local Similarity 100.0%; Pred. No. 2,8e-112;

Matches 378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TACGCTCTTAAGACGACTTAAGGGGAGATACGACGCGCTTTAGAGAAATATGG 60

DB 1 TACGCTCTTAAGACGACTTAAGGGGAGATACGACGCGCTTTAGAGAAATATGG 60

QY 61 AAAACACCCAGCGAGTCTCTCAAGCTGATGTGTCTTGAGAGTGAAGGATGC 120

DB 61 AAAACACCCAGCGAGTCTCTCAAGCTGATGTGTCTTGAGAGTGAAGGATGC 120

QY 121 ACGGTTGTTAAGCCCTGTCTTCCGTTTATCTAATCTAATCTTCTTGAATAAAG 180

DB 121 ACGGTTGTTAAGCCCTGTCTTCCGTTTATCTAATCTAATCTTCTTGAATAAAG 180

QY 131 CTCCTGCCAGTACTGCTTTTATGCTCAACATGCTTTGACTGTTGAAAAGACC 240

DB 131 CTCCTGCCAGTACTGCTTTTATGCTCAACATGCTTTGACTGTTGAAAAGACC 240

QY 181 CTCCTGCCAGTACTGCTTTTATGCTCAACATGCTTTGACTGTTGAAAAGACC 240

DB 241 TTGGCACAATGTAAGGATGATGATGATGATGATGATGATGATGATGATGATG 300

QY 241 TTGGCACAATGTAAGGATGATGATGATGATGATGATGATGATGATGATGATG 300

QY 301 CTATGTTGTAAGTATGAGAGAGTCTTCTTGGCAAAAGATCTCCCGGAAGAGATG 360

DB 301 CTATGTTGTAAGTATGAGAGAGTCTTCTTGGCAAAAGATCTCCCGGAAGAGATG 360

QY 361 CCATTGGAGAACCATGCG 378

DB 361 CCATTGGAGAACCATGCG 378

RESULT 2
AAS37218
ID AAS37218 standard; cDNA; 327 BP.

XX AAS37218;

XX 17-DEC-2001 (first entry)

XX Novel human diagnostic and therapeutic gene #276.

XX Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.

XX Homo sapiens.

XX WO20016753-A2.

XX 13-SEP-2001.

XX 09-MAR-2001; 2001WO-US07787.

XX 09-MAR-2000; 2000US-0188609.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

XX Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;

XX Drmanac R, Cervenjakov R, Dickson M, Drmanac S, Labat I;

XX Leshkowitz D, Kita D, Garcia V, Jones WL, Steche-Crain B;

XX WPI; 2001-530177/58.

XX New polynucleotides and polypeptides, useful for diagnosis and
treatment of breast, lung and colon cancer.

PS Claim 1; Page 665; 1193pp; English.

XX The invention relates to new polynucleotides and polypeptides, useful for
PS diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.

XX Sequence 327 BP; 92 A; 64 C; 81 G; 90 T; 0 other;

SO Query Match 79.3%; Score 299.6; DB 22; Length 327;
Best Local Similarity 98.7%; Pred. No. 6,8e-87;

Matches 302; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 28 GAATGACGAGCGGCTCTTGAAGAACATATGAAACACCCAGCGGAGTCTCTACA 87

DB 21 GAATGACGAGCGGCTCTTGAAGAACATATGAAACACCCAGCGGAGTCTCTACA 80

QY 88 AGCTTGAATGTGTCTCTGAGCTGAAGATGACGAGTGTAAAGCCCTGTTCTTCC 147

DB 81 AGCTTGAATGTGTCTCTGAGCTGAAGATGACGAGTGTAAAGCCCTGTTCTTCC 140

QY 148 GTTGTATTAATCTAATGTTCTTGGATTAACCTCCCTGCCAAGTACTGTTTTC 207

DB 141 GTTGTATTAATCTAATGTTCTTGGATTAACCTCCCTGCCAAGTACTGTTTTC 200

QY 208 TGCTCAACATGCTTGTGCTGTTGAAAGAGACCTTGGCACACATTTGAAGGATGGAT 267

DB 201 TGCTCAACATGCTTGTGCTGTTGAAAGAGACCTTGGCACACATTTGAAGGATGGAT 260

QY 268 GGAGATGCCAATCCATGATGATGATGATGATGATGATGATGATGATGATGATG 327

DB 261 GGAGATGCCAATCCATGATGATGATGATGATGATGATGATGATGATGATGATG 320

QY 328 TCTTGG 333

DB 321 TCTTGG 326

RESULT 3
AAS37237
ID AAS37237 standard; cDNA; 322 BP.

XX AAS37237;

XX 17-DEC-2001 (first entry)

XX Novel human diagnostic and therapeutic gene #295.

XX Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.

XX Homo sapiens.

XX WO20016753-A2.

XX 13-SEP-2001.

XX (CHIR) CHIRON CORP.

XX 09-MAR-2001: 200IMO-US07787.
PE
XX
PR 09-MAR-2000: 200OUS-0188609.
XX
PA (CHIR) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
XX Reinhard C, Randazzo F, Kennedy G, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crivenjaevak R, Dickson M, Drmanac S, Labat I;
XX Leshkowitz D, Kita D, Garcia V, Jones WL, Stachne-Crain B;
DR WPI, 2001-530177/58.
XX
PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer -
XX
PS Claim 1, Page 671: 1193pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX
SQ Sequence 318 BP: 90 A; 63 C; 76 G; 89 T; 0 other;
XX

Query Match 72.8%; Score 275; DB 22; Length 318;
Best Local Similarity 98.2%; Pred. No. 6.3e-79;
Matches 278; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 GAATACGAGAGGGGCTTAAAGAGAAACATATGAAAGAAACCCAGCGAGTCTTCACA 87
DB 36 GAATACGAGAGGAGCTCTGAGAGAACTAAAGAAACACCCAGCGAGTCTTCACA 95
QY 88 AGCTTGAATGCTGCTTCTGAGAGCTGAAGATGCACGGTGTAAAGCCCGCTGTTTCC 147
DB 96 AGCTTGAATGCTGCTTCTGAGAGCTGAAGATGCACGGTGTAAAGCCCGCTGTTTCC 155
QY 148 GTTGTTTAATCTAATGTTCTTTGGAAATAAAACCTCCCTGCCAAGTAAGTACTGTTTGA 207
DB 156 GTTGTTTAATCTAATGTTCTTTGGAAATAAAACCTCCCTGCCAAGTAAGTACTGTTTGA 215
QY 208 TGTCTCAACATGCTTGAAGTGTGTAAGAAAGACGCTTGGACACACATGAAGAGATGCTAT 267
DB 216 TGTCTCAACATGCTTGAAGTGTGTAAGAAAGACGCTTGGACACACATGAAGAGATGCTAT 275
QY 268 GGAAGATGCCAATCCATGGAATCAGGTGGCGCAGCTATGTTGGT 310
DB 276 GGAAGATGCCAATCCATGGAATCAGGTGGCGCAGCTATGTTGGT 318

RESULT 5
ABL33878
ID ABL33878 standard; DNA: 5378 BP.
AC ABL33878;
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1851.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KM antiarteriosclerotic; antianaemic; cytosstatic; noctopic;
KW neuroprotective; anti-HIV; anticomvulsant; ophthalmological;
KM antineumatic; antiarthritis; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200200928-A2.
 PD
 XX
 PD 03-JAN-2002.
 PF 02-JUL-2001; 2001MO-EP07537.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 PA (EPIC-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 PS
 XX Claim 1; SEQ ID NO 1851; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 SQ Sequence 5378 BP; 1499 A; 134 C; 1201 G; 2544 T; 0 other;
 Query Match 8.9%; Score 33.8; DB 24; Length 5378;
 Best Local Similarity 48.2%; Pred. No. 3.1;
 Matches 95; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
 QY 137 TGTCTCTTCCGTTGTTAATCAATGTTCTTGGAAATAAACCTCCGCAAGTAGT 196
 DB 1377 TTTTAAATTTAGTGAATTTATTTATTTTGAATTTTATTAATATTTAGTGT 1436
 QY 197 ACTTGTCTTATGCTCAACATGCTTGACCTTGAGAAAGAGACTTTGGACACATTGAA 236
 DB 1437 ACGAAGTTTAGTCAATTTATTTGCGATCGTTGGAGAAAGAGCTTTTAAATATTAG 1496
 QY 257 GGGATGCTATGAGATGCAATCATCGAATGAGGCGGCGAGCTATGTTGCTAT 316
 DB 1497 ATTAATTTTATTTAATTTATTTTATTTAGGGGTGAAGTAAAGTTTGTGTTAAAGGAT 1556
 QY 317 AGCAGAACTCTCTTGG 333
 DB 1557 AATTGAGTATTATTAG 1573
 RESULT 6
 ABA08523
 ID ABA08523 standard; CDNA; 492 BP.
 XX
 AC ABA08523;
 XX
 DT 11-JAN-2002 (first entry)
 XX
 DE Human secreted protein homologue-encoding cDNA, SEQ ID NO:299.
 XX
 KW Human; cytokines; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemokinesis; chemokinesis; thrombolytic; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;

KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; antiinflammatory;
 KW antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;
 KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 KW antifungal; vulnerrary; antitumor; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200157188-A2.
 PD
 XX
 PD 09-AUG-2001.
 PF 05-FEB-2001; 2001MO-US03800.
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PA (HYSE-) HYSEO INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 DR WPI; 2001-457740/49.
 XX
 DR P-PSDB; ABB11279.
 XX
 PT Human proteins and DNA encoding sequences useful for preventing,
 PT treating or ameliorating a medical condition in a mammalian subject
 PT e.g. arthritis and cancer -
 PS
 XX Claim 1; Page 440; 1963pp; English.
 XX
 CC Sequences ABA10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides
 CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the
 CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoiesis regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness,
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a cDNA encoding a
 CC novel human polypeptide of the invention.
 XX
 SQ Sequence 492 BP; 124 A; 97 C; 99 G; 172 T; 0 other;

Query Match 8.9%; Score 33.6; DB 22; Length 492;
 Best Local Similarity 52.9%; Pred. No. 1.3;
 Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

DB 91 TTGAATGTCGTTCTGAGCTGAGATGACGAGCTGTTAAAGCCCGTCTTTCCGCT 150
 51 TGAACCTTACGCTTCTGTTGGTTGGAGTCATCAGTACCTGGGTATTTCCAGGC 110
 OY 151 GTTAACTCTAATGTTCTTGGATTAATAAACCTCCCTGCCAAGTAGTACTGTTTATGC 210
 DB 111 ATATATGTCGATTTTAAACATGACGCTCCTACATGCCAAAGTTCTTGGTTTGTGC 170
 OY 211 TCAACATGCTTGACCT 226
 DB 171 TTGCTGCTGATGATT 186

RESULT 7 AAC29807/c

ID AAC29807 standard; cDNA; 158 BP.

AC 29807;

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 33882.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GIST) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 diagnostic, forensic, gene therapy and chromosome mapping procedures -

Claim 1; SEQ ID 33882; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from
 mRNAs encoding secreted proteins. No ORF has yet been conclusively
 identified within the present sequence. The 5' ESTs were prepared from
 total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 sequences usually correspond mainly to the 3' untranslated region (UTR)
 of the mRNA because they are often obtained from oligo-dT primed cDNA
 libraries. Such ESTs are not well suited for isolating cDNA sequences
 derived from the 5' ends of mRNAs and even in those cases where longer
 cDNA sequences have been obtained, the full 5' UTR is rarely included.
 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 They are used to obtain upstream regulatory sequences and to design
 expression and secretion vectors.

Sequence 158 BP; 35 A; 34 C; 34 G; 55 T; 0 other;

Query Match 8.8%; Score 33.4; DB 21; Length 158;
 Best Local Similarity 53.4%; Pred. No. 0.9;
 Matches 70; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 1 TACGCTGCTTAAACACACTTAAAGGGGAATGACGACGCGCTTTAGAGCAATATGG 60
 DB 144 TATAGCTCCGAGAGAGCAACAAGGAATCTCATTTCTCTCTCTACAGAGA 85
 OY 61 AAAACACCAAGCCGAGTCTTCACAAAGCTTGAATGTGTTTGGAGCTGAAGATGC 120
 DB 84 ACAGCATTAATCCATGAGCCACTAACAGCTCTGGAAGATGAATGAGTGTGATGAAGA 25
 OY 121 ACGGTGTAA 131
 DB 24 AGGGAGAGAAA 14

RESULT 8 AA24099/c

ID AA24099 standard; DNA; 240 BP.

AA24099;

04-FEB-2000 (first entry)

Rat NF-1 DNA fragment 2.

NF-1; rat; primer; mutagenicity; mutation rate; polymerase error rate;
 radiation; ss.

Rattus sp.

DE19818422-A1.

28-OCT-1999.

24-APR-1998; 98DE-1018422.

24-APR-1998; 98DE-1018422.

(HARJ) HARGES S.

Harges S;

WPI; 1999-602407/52.

Method for determining absolute mutagenicity and error rate of
 polymerases in amplification reactions -

Examples; Fig 5; 18pp; German.

This invention describes a novel method for determining the absolute
 mutagenicity of a substance. The method is used to determine the
 mutation rate associated with particular substances and the error rates
 for polymerases during amplification reactions. It is therefore useful
 for monitoring nutrients or medicines and environmental factors such as
 radiation from televisions or computer screens. The method is applicable
 to all types of mutations (collectively or individually), is unaffected
 by repair mechanisms or in vivo selection mechanisms, and is an in vitro
 process so animal testing is not required. AA24098-24105 represent
 rat NF-1 DNA fragments used to illustrate the method of the invention.

Sequence 240 BP; 66 A; 53 C; 53 G; 68 T; 0 other;

Query Match 8.8%; Score 33.4; DB 20; Length 240;
 Best Local Similarity 72.9%; Pred. No. 1.1;
 Matches 43; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 148 GTTCTTAATCTAATGTTCTTTGGAATAAAACCTCCGCCAAGTAGTGTGTTT 206
 DB 60 GTGGACATCTGATGCTATATCTAAGAAAACCTCGGCCAAGTAGTGTGTTT 2

RESULT 9 AA24100/c

ID AA24100 standard; DNA; 240 BP.

AC AA224100;
 XX
 DT 04-FEB-2000 (first entry)
 XX
 DE Rat NF-1 DNA fragment 3.
 XX
 KW NF-1; rat; primer; mutagenicity; mutation rate; polymerase error rate;
 KM radiation; ss.
 XX
 OS Rattus sp.
 XX
 PN DE19918422-A1.
 XX
 PD 28-OCT-1999.
 XX
 PF 24-APR-1998; 98DE-1018422.
 XX
 PR 24-APR-1998; 98DE-1018422.
 XX
 PA (HARD/) HARJES S.
 XX
 PI Harjes S;
 XX
 DR WPI; 1999-602407/52.
 XX
 PT Method for determining absolute mutagenicity and error rate of
 PT polymerases in amplification reactions -
 XX
 PS Examples; Fig 5; 18bp; German.
 XX
 CC This invention describes a novel method for determining the absolute
 CC mutagenicity of a substance. The method is used to determine the
 CC mutation rate associated with particular substances and the error rates
 CC for polymerases during amplification reactions. It is therefore useful
 CC for monitoring nutrients or medicines and environmental factors such as
 CC radiation from televisions or computer screens. The method is applicable
 CC to all types of mutations (collectively or individually), is unaffected
 CC by repair mechanisms or in vivo selection mechanisms, and is an in vitro
 CC process so animal testing is not required. AA224098-224105 represent
 CC rat NF-1 DNA fragments used to illustrate the method of the invention.
 CC
 SQ Sequence 240 BP; 66 A; 53 C; 53 G; 68 T; 0 other;

Query Match 8.8%; Score 33.4; DB 20; Length 240;
 Best Local Similarity 72.9%; Pred. No. 1.1;
 Matches 43; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 148 GTTGTAACTCAATGTTCTTTGGAAATAAAACCTCCGCAAGTAGTACTTGGTTT 206
 DB 60 GTGGACAATCTGATCTATATCTAAGAAAACCTCCGTCGCAAGTCGAGTGTCTTT 2

RESULT 10
 ID AAAB1467/c
 AAAB1467 standard; DNA; 25509 BP.
 AC
 AC AAAB1467;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_15 SEQ ID NO:15.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenics;
 KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KM Meningococcus B; MenB; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO200022430-A2.
 XX
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999; 99WO-US23573.

XX
 PR 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Frizer CM, Hickey E, Peterson J, Tettein H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 XX
 DR WPI; 2000-318079/27.
 XX
 PT Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 XX
 PS Claim 7; Page 399-406; 1760pp; English.

CC The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAAB1453 to AAAB2414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAAB1260 to AAAB1303 and AAAB25620 to AAAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAAB1254 to
 CC AAAB1259 and AAAB1304 to AAAB1321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAAB1322 to
 CC AAAB1452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 CC
 SQ Sequence 25509 BP; 6157 A; 6156 C; 6549 G; 6647 T; 0 other;

Query Match 8.8%; Score 33.4; DB 21; Length 25509;
 Best Local Similarity 52.5%; Pred. No. 8.4;
 Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 169 TGGATAAAAAACCTCCCTGCCAAGTAGTACTTGTATATCTCAACATGCTTGACGT 228
 DB 23305 TGATATATAAAATCATTCAGCAAGAAACATTAATATTTTGAATATATGCTCCCTTAATAT 23246

QY 229 TGAAGAAGACCTTTGGACACATTTGAAGGATGGTATGAGATGCCAATCCATCGAAT 288
 DB 23245 CGAAGAGAAAGATCTCGAAAAATATATATAAAGCGGTGTTCCGCAAAATATAGTTT 23186

QY 289 CAGGTGGCGCAGCTATGTT 307
 DB 23185 TAATTGGACAGCTATATT 23167

RESULT 11
 ID AAF21607/c
 AAF21607 standard; DNA; 349980 BP.
 AC
 AC AAF21607;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.

PR 14-SEP-2000; 2000US-0232398-
 PR 14-SEP-2000; 2000US-0232399-
 PR 14-SEP-2000; 2000US-0232400-
 PR 14-SEP-2000; 2000US-0232401-
 PR 14-SEP-2000; 2000US-0233063-
 PR 14-SEP-2000; 2000US-0233064-
 PR 14-SEP-2000; 2000US-0233065-
 PR 21-SEP-2000; 2000US-0234223-
 PR 21-SEP-2000; 2000US-0234274-
 PR 25-SEP-2000; 2000US-0234997-
 PR 25-SEP-2000; 2000US-0234998-
 PR 26-SEP-2000; 2000US-0235834-
 PR 27-SEP-2000; 2000US-0235834-
 PR 27-SEP-2000; 2000US-0235836-
 PR 29-SEP-2000; 2000US-0236327-
 PR 29-SEP-2000; 2000US-0236367-
 PR 29-SEP-2000; 2000US-0236368-
 PR 29-SEP-2000; 2000US-0236369-
 PR 29-SEP-2000; 2000US-0236370-
 PR 02-OCT-2000; 2000US-0236802-
 PR 02-OCT-2000; 2000US-0237037-
 PR 02-OCT-2000; 2000US-0237038-
 PR 02-OCT-2000; 2000US-0237039-
 PR 02-OCT-2000; 2000US-0237040-
 PR 13-OCT-2000; 2000US-0239935-
 PR 13-OCT-2000; 2000US-0239937-
 PR 20-OCT-2000; 2000US-0240960-
 PR 20-OCT-2000; 2000US-0241221-
 PR 20-OCT-2000; 2000US-0241785-
 PR 20-OCT-2000; 2000US-0241786-
 PR 20-OCT-2000; 2000US-0241787-
 PR 20-OCT-2000; 2000US-0241808-
 PR 20-OCT-2000; 2000US-0241809-
 PR 20-OCT-2000; 2000US-0241826-
 PR 01-NOV-2000; 2000US-0244617-
 PR 08-NOV-2000; 2000US-0246474-
 PR 08-NOV-2000; 2000US-0246475-
 PR 08-NOV-2000; 2000US-0246476-
 PR 08-NOV-2000; 2000US-0246477-
 PR 08-NOV-2000; 2000US-0246478-
 PR 08-NOV-2000; 2000US-0246523-
 PR 08-NOV-2000; 2000US-0246524-
 PR 08-NOV-2000; 2000US-0246525-
 PR 08-NOV-2000; 2000US-0246526-
 PR 08-NOV-2000; 2000US-0246527-
 PR 08-NOV-2000; 2000US-0246528-
 PR 08-NOV-2000; 2000US-0246532-
 PR 08-NOV-2000; 2000US-0246609-
 PR 08-NOV-2000; 2000US-0246610-
 PR 08-NOV-2000; 2000US-0246611-
 PR 08-NOV-2000; 2000US-0246613-
 PR 17-NOV-2000; 2000US-0249207-
 PR 17-NOV-2000; 2000US-0249208-
 PR 17-NOV-2000; 2000US-0249209-
 PR 17-NOV-2000; 2000US-0249210-
 PR 17-NOV-2000; 2000US-0249211-
 PR 17-NOV-2000; 2000US-0249212-
 PR 17-NOV-2000; 2000US-0249213-
 PR 17-NOV-2000; 2000US-0249214-
 PR 17-NOV-2000; 2000US-0249215-
 PR 17-NOV-2000; 2000US-0249216-
 PR 17-NOV-2000; 2000US-0249217-
 PR 17-NOV-2000; 2000US-0249218-
 PR 17-NOV-2000; 2000US-0249244-
 PR 17-NOV-2000; 2000US-0249245-
 PR 17-NOV-2000; 2000US-0249264-
 PR 17-NOV-2000; 2000US-0249265-
 PR 17-NOV-2000; 2000US-0249297-
 PR 17-NOV-2000; 2000US-0249299-
 PR 17-NOV-2000; 2000US-0249300-
 PR 01-DEC-2000; 2000US-0250160-
 PR 01-DEC-2000; 2000US-0250391-
 PR 05-DEC-2000; 2000US-0251030-

PR 05-DEC-2000; 2000US-0251988-
 PR 05-DEC-2000; 2000US-0256719-
 PR 06-DEC-2000; 2000US-0251479-
 PR 08-DEC-2000; 2000US-0251856-
 PR 08-DEC-2000; 2000US-0251868-
 PR 08-DEC-2000; 2000US-0251869-
 PR 08-DEC-2000; 2000US-0251989-
 PR 08-DEC-2000; 2000US-0251990-
 PR 11-DEC-2000; 2000US-0254097-
 PR 05-JAN-2001; 2001US-0259678-
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI: 2001-483426/52.
 DR P-PSDB; AAM84888.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 PS
 XX Claim 1: SEQ ID NO 2729; 3071bp + Sequence Listing; English.
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 303 BP; 100 A; 62 C; 45 G; 94 T; 2 other;
 XX
 Query Match 8.7%; Score 33; DB 22; Length 303;
 Best Local Similarity 48.1%; Pred. NO. 1.6;
 Matches 90; Conservative 1; Mismatches 96; Indels 0; Gaps 0;
 QY 20 TTAAAGGGGAATGACGAGCGCTCTTAAGACATATGAAAAACCCAAAGCCGAGT 79
 DB 297 TTGGGTGGGGAGCTGCCCAAGTCAGAAATGCAATTAGAAAAACAATGTACGAACC 238
 QY 80 CTCACACAGCTGAATGTGTCTGAGCTGAAGATGACAGCTTTAGACCCCTGT 139
 DB 237 CTCCTCACACATGATTTCAATGTCTCTGCCCAATCCTGTGAATATATATAGCTCTCT 178
 QY 140 TCTTTCCGTTGTTTATCTAATCTTTGGAATAAACCTCCCTGCCAAGTAGTACT 199
 DB 177 TGTTCACAGGGGTTTCGTGAATCTCTTTAAGTTAATGATGAAATGACAAATGTTCT 118
 QY 200 TGGTTT 206
 DB 117 GGTCTT 111
 RESULT 13
 AAZ24101/c
 ID AAZ24101 standard; DNA: 240 BP.
 XX
 AC AAZ24101;
 XX
 DT 04-FEB-2000 (first entry)
 XX

DE Rat NF-1 DNA fragment 4.
 XX
 XX NF-1; rat; primer; mutagenicity; mutation rate; polymerase error rate;
 KW radiation; ss.
 XX
 OS Rattus sp.
 XX
 PN DE19818422-A1.
 XX
 PD 28-OCT-1999.
 XX
 PF 24-APR-1998; 98DE-1018422.
 XX
 PR 24-APR-1998; 98DE-1018422.
 XX
 PA (HARJ/) HARJES S.
 XX
 PI Harjes S;
 XX
 DR WPI; 1999-602407/52.
 XX
 PT Method for determining absolute mutagenicity and error rate of
 PT polymerases in amplification reactions -
 PS Examples; Fig 5; 18pp; German.
 XX
 CC This invention describes a novel method for determining the absolute
 CC mutagenicity of a substance. The method is used to determine the
 CC mutation rate associated with particular substances and the error rates
 CC for polymerases during amplification reactions. It is therefore useful
 CC for monitoring nutrients or medicines and environmental factors such as
 CC radiation from televisions or computer screens. The method is applicable
 CC to all types of mutations (collectively or individually), is unaffected
 CC by repair mechanisms or in vivo selection mechanisms, and is an in vitro
 CC process so animal testing is not required. AA224098-224105 represent
 CC rat NF-1 DNA fragments used to illustrate the method of the invention.
 XX
 SQ Sequence 240 BP; 66 A; 52 C; 54 G; 68 T; 0 other;
 XX

Query Match 8.7%; Score 32.8; DB 20; Length 240;
 Best Local Similarity 76.9%; Pred. No. 1.7;
 Matches 40; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 155 AATCTAATGTTCTTGGAAATAAAACCTCCCTGCCAGTAGTACTGCTTTT 206
 Db 53 AATCTGATGCTATATCTAAGAAAAACCTCCCTGCCAAGTCGAGTTCCTTT 2

RESULT 14
 AA224102/c
 ID AA224102 standard; DNA; 240 BP.
 XX
 AC AA224102;
 XX
 DT 04-FEB-2000 (first entry)
 XX
 DE Rat NF-1 DNA fragment 5.
 XX
 XX NF-1; rat; primer; mutagenicity; mutation rate; polymerase error rate;
 KW radiation; ss.
 XX
 OS Rattus sp.
 XX
 PN DE19818422-A1.
 XX
 PD 28-OCT-1999.
 XX
 PF 24-APR-1998; 98DE-1018422.
 XX
 PR 24-APR-1998; 98DE-1018422.
 XX
 PA (HARJ/) HARJES S.
 XX

PI Harjes S;
 XX
 XX WPI; 1999-602407/52.
 DR
 XX
 XX Method for determining absolute mutagenicity and error rate of
 PT polymerases in amplification reactions -
 PS Examples; Fig 5; 18pp; German.
 XX
 CC This invention describes a novel method for determining the absolute
 CC mutagenicity of a substance. The method is used to determine the
 CC mutation rate associated with particular substances and the error rates
 CC for polymerases during amplification reactions. It is therefore useful
 CC for monitoring nutrients or medicines and environmental factors such as
 CC radiation from televisions or computer screens. The method is applicable
 CC to all types of mutations (collectively or individually), is unaffected
 CC by repair mechanisms or in vivo selection mechanisms, and is an in vitro
 CC process so animal testing is not required. AA224098-224105 represent
 CC rat NF-1 DNA fragments used to illustrate the method of the invention.
 XX
 SQ Sequence 240 BP; 66 A; 52 C; 54 G; 68 T; 0 other;
 XX

Query Match 8.7%; Score 32.8; DB 20; Length 240;
 Best Local Similarity 76.9%; Pred. No. 1.7;
 Matches 40; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 155 AATCTAATGTTCTTGGAAATAAAACCTCCCTGCCAGTAGTACTGCTTTT 206
 Db 53 AATCTGATGCTATATCTAAGAAAAACCTCCCTGCCAAGTCGAGTTCCTTT 2

RESULT 15
 AA224103/c
 ID AA224103 standard; DNA; 240 BP.
 XX
 AC AA224103;
 XX
 DT 04-FEB-2000 (first entry)
 XX
 DE Rat NF-1 DNA fragment 6.
 XX
 XX NF-1; rat; primer; mutagenicity; mutation rate; polymerase error rate;
 KW radiation; ss.
 XX
 OS Rattus sp.
 XX
 PN DE19818422-A1.
 XX
 PD 28-OCT-1999.
 XX
 PF 24-APR-1998; 98DE-1018422.
 XX
 PR 24-APR-1998; 98DE-1018422.
 XX
 PA (HARJ/) HARJES S.
 XX
 PI Harjes S;
 XX
 DR WPI; 1999-602407/52.
 XX
 PT Method for determining absolute mutagenicity and error rate of
 PT polymerases in amplification reactions -
 PS Examples; Fig 5; 18pp; German.
 XX
 CC This invention describes a novel method for determining the absolute
 CC mutagenicity of a substance. The method is used to determine the
 CC mutation rate associated with particular substances and the error rates
 CC for polymerases during amplification reactions. It is therefore useful
 CC for monitoring nutrients or medicines and environmental factors such as
 CC radiation from televisions or computer screens. The method is applicable
 CC to all types of mutations (collectively or individually), is unaffected
 CC by repair mechanisms or in vivo selection mechanisms, and is an in vitro

CC process so animal testing is not required. AA24098-224105 represent
 CC rat NF-1 DNA fragments used to illustrate the method of the invention.
 XX
 SO Sequence 240 BP; 66 A; 52 C; 54 G; 68 T; 0 other;

Query Match 8.7%; Score 32.8; DB 20; Length 240;
 Best Local Similarity 76.9%; Pred. No. 1.7;
 Matches 40; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 155 ATCTATGTCTTCTTGGATAAAACCTCCCTGCCAAGTACTGTTGTTT 206
 ||||| ||| | | | ||||| ||||| ||||| ||||| |||||
 Db 53 AATCTGATGCTATATCTTAAGAAAAACCTCCGTCGCAAGTCGAAGTCTTTT 2

Search completed: January 31, 2003, 01:17:24
 Job time : 514 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2003, 01:15:03 : Search time 2759 seconds

(without alignments)
3987.264 Million cell updates/sec

Title: US-09-803-719-222

Perfect score: 378

Sequence: 1 taaggctgcttaagacgact.....taacattggaagaacatgcg 378

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sta:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sta:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	378	100.0	378	6 AX245292	AX245292 Sequence
2	299.6	79.3	327	6 AX245346	AX245346 Sequence
3	279.6	74.0	322	6 AX245365	AX245365 Sequence
4	275	72.8	318	6 AX245370	AX245370 Sequence
5	119	31.5	59231	9 AL135937	AL135937 Human DNA
6	67.8	17.9	162394	2 AC139775	AC139775 Rattus no
7	53.8	14.2	223604	10 AL713981	AL713981 Mouse DNA
8	41	10.8	176543	2 AL845312	AL845312 Danio rer
9	39.4	10.4	176030	2 AC129650	AC129650 Rattus no
10	39.2	10.4	91071	8 AC002339	AC002339 Arabidops
11	38.4	10.2	130206	9 AC104652	AC104652 Homo sapi
12	38.4	10.2	152822	9 AC108866	AC108866 Homo sapi
13	38.4	10.2	150912	2 AC080170	AC080170 Homo sapi
14	38.2	10.1	30960	3 AC024785	AC024785 Caenorhab
15	37.8	10.0	174196	2 AL732628	AL732628 Danio rer
16	37.8	10.0	177368	2 AC097813	AC097813 Rattus no
17	37.4	9.9	137193	2 AC110131	AC110131 Rattus no
18	37.4	9.9	160582	2 AC108063	AC108063 Homo sapi
19	37.4	9.9	192297	2 AC129172	AC129172 Rattus no
20	37.4	9.9	259272	2 AC096436	AC096436 Rattus no
21	37.2	9.8	165118	2 AC107535	AC107535 Rattus no
22	37	9.8	1350	8 NMU91924	U91924 Nicotiana t
23	37	9.8	215188	2 AC124196	AC124196 Mus muscu
24	36.4	9.6	158968	10 AL662785	AL662785 Mouse DNA
25	36.2	9.6	1980	3 LOU59830	LOU59830 Loligo opal
26	36.2	9.6	50840	9 AC093598	AC093598 Homo sapi
27	36.2	9.6	127508	2 AP002964	AP002964 Homo sapi
28	36.2	9.6	162921	2 AC013528	AC013528 Homo sapi
29	36.2	9.6	173526	2 AL160036	AL160036 Human DNA
30	36	9.5	165282	2 AP000822	AP000822 Homo sapi
31	35.8	9.5	62254	2 AC100028	AC100028 Mus muscu
32	35.8	9.5	62582	2 AC103992	AC103992 Homo sapi
33	35.8	9.5	162605	2 AC068698	AC068698 Homo sapi
34	35.8	9.5	173760	2 AL354653	AL354653 Human DNA
35	35.8	9.5	188515	2 AC102531	AC102531 Mus muscu
36	35.8	9.5	210386	10 AL663065	AL663065 Mouse DNA
37	35.6	9.4	110000	2 AC092450	AC092450 Contination (3 of
38	35.6	9.4	117296	9 AC092865	AC092865 Homo sapi
39	35.6	9.4	155415	2 AC108886	AC108886 Rattus no
40	35.6	9.4	162173	8 AP002953	AP002953 Oryza sat
41	35.6	9.4	163317	9 AC096564	AC096564 Homo sapi
42	35.6	9.4	204632	2 AC097583	AC097583 Rattus no
43	35.4	9.4	1537	8 AB015609	AB015609 Nicotiana
44	35.4	9.4	2325	8 AF033100	AF033100 Nicotiana
45	35.4	9.4	7005	1 AB010426	AB010426 Phytolas

ALIGNMENTS

```

RESULT 1
AX245292 378 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 222 from Patent WO0166753.
ACCESSION AX245292
VERSION AX245292.1 GI:15859966
KEYWORDS
SOURCE human.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 378)
Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
Kessam,A., Lamson,G., Drmanac,R., Ckvenjakov,R., Dickson,M.,

```

	Best Local Similarity	98.7%;	Pred. No. 1.2e-79;	Mismatches	4;	Indels	0;	Gaps	0;
	Matches	302;	Conservative	0;					
OY	28	GAATGACCGCAGCGGGCTTTAGAGAAACATATGGAATAAACCCAGCCGAGTCTTCACA	87						
Db	21	GAATGACCGCAGAGCTGTGAGAGAAACATTAAGAAAACACCAGCCGAGTCTTCACA	80						
OY	88	AGCTTGAAATGTGTGTCTTGCAGACTGAAGATGACGCTTGAAGCCCCGTCTTCTCC	147						
Db	81	AGCTTGAAATGTGTGTCTTGCAGACTGAAGATGACGCTTGAAGCCCCGTCTTCTCC	140						
OY	148	GTTGTTTAATCATATGTCTTGTGAAATAAAAACCCCTGCCAAGTAGTACTGTGTTTTA	207						
Db	141	GTTGTTTAATCATATGTCTTGTGAAATAAAAACCCCTGCCAAGTAGTACTGTGTTTTA	200						
OY	208	TGCTCAACATCGCTTGTGACTGTGAAAAAGACACTTTGGCACACATTGGAAGGATGTGAT	267						
Db	201	TGCTCAACATCGCTTGTGACTGTGAAAAAGACACTTTGGCACACATTGGAAGGATGTGAT	260						
OY	268	GGAGATGCCCATCCATGGAATACAGGTGGCGCAGCATATGTTGGTAGCTTATGACAAGTCT	327						
Db	261	GGAGATGCCCATCCATGGAATACAGGTGGCGCAGCATATGTTGGTAGCTTATGACAAGTCT	320						
OY	328	TCTTGG 333							
Db	321	TCTTGG 326							
RESULT 3									
LOCUS	AX245365	322 bp	DNA	linear	PAT 28-SEP-2001				
DEFINITION	Sequence 295 from Patent WO0166753.								
ACCESSION	AX245365								
VERSION	AX245365.1	GI:15860039							
KEYWORDS									
SOURCE	human.								
ORGANISM	Homo sapiens								
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.								
REFERENCE	1 (bases 1 to 322)								
AUTHORS	Williams, L.T., Escobedo, J., Innis, M.A., Garcia, P.D.,								
	Sudduth-Klinger, J., Reinhard, C., Randazzo, F., Kennedy, G.C., Pot, D.,								
	Kassam, A., Lamson, G., Drmanac, R., Crkvenjakov, R., Dickson, M.,								
	Drmanac, S., Labat, I., Leschowitz, D., Kita, D., Garcia, V. and								
	Stache-Cralin, B.								
TITLE	Human genes and gene expression products								
JOURNAL	Patent: WO 0166753-A 295 13-SEP-2001;								
FEATURES	Chiron Corporation (US); Hyseq Inc. (US)								
source	location/Qualifiers								
	1..322								
	/organism="Homo sapiens"								
BASE COUNT	90 a 64 c 78 g 89 t 1 others								
ORIGIN									
	Query Match 74.0%; Score 279.6; DB 6; Length 322;								
	Best Local Similarity 98.6%; Pred. No. 1.4e-73;								
	Matches 282; Conservative 0; Mismatches 4; Indels 0; Gaps 0;								
OY	28	GAATGACCGCAGCGCTCTTGAAGAATATGAAAAACCCAGCCGAGTCTTCACA	87						
Db	36	GAATGACCGCAGCAGCTCTGAGAGAAACATTAAGAAAACACCAGCCGAGTCTTCACA	95						
OY	88	AGCTTGAAATGTGTGTCTTGCAGACTGAAGATGACGCTTGAAGCCCCGTCTTCTCC	147						
Db	96	AGCTTGAAATGTGTGTCTTGCAGACTGAAGATGACGCTTGAAGCCCCGTCTTCTCC	155						
OY	148	GTTGTTTAATCATATGTCTTGTGAAATAAAAACCCCTGCCAAGTAGTACTGTGTTTTA	207						
Db	156	GTTGTTTAATCATATGTCTTGTGAAATAAAAACCCCTGCCAAGTAGTACTGTGTTTTA	215						
OY	208	TGCTCAACATCGCTTGTGACTGTGAAAAAGACACTTTGGCACACATTGGAAGGATGTGAT	267						

DB 216 TGCTCAACATGCTTTGACCTGTGAAAAGACCTTTGGCACACATTCGAGGATGCTGAT 275
 QY 268 GGAGATGCCATCCATGGAATCAGTGGCGCCAGCTATGTTGGTACG 313
 DB 276 GGAGATGCCATCCATGGAATCAGTGGCGCACAGCTATGTTGGTACG 321

RESULT 4
 AX245370 318 bp DNA linear PAT 28-SEP-2001
 LOCUS Sequence 300 from Patent WO0166753.
 DEFINITION AX245370
 ACCESSION AX245370
 VERSION AX245370.1 GI:15860044
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 318)
 AUTHORS Williams, L.T., Escobedo, J., Innis, M.A., Garcia, P.D., Suduth-Klinger, J., Reinhard, C., Randazzo, F., Kennedy, G.C., Pot, D., Kassam, A., Lamson, G., Drmanac, R., Crkvenjakov, R., Dickson, M., Drmanac, S., Labat, I., Leshkowitz, D., Kita, D., Garcia, V. and Stache-Crain, B.
 Human genes and gene expression products
 Patent: WO 0166753-A 300 13-SEP-2001;
 Chiron Corporation (US) ; Hyseq Inc. (US)
 FEATURES
 source 1..318
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 90 a 63 c 76 g 89 t
 ORIGIN

Query Match 72.8% Score 275; DB 6; Length 318;
 Best Local Similarity 98.2% Pred. No. 3.4e-72;
 Matches 278; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 GAATGACGCGCGCTCTTAGAGAACATATGAAAAACACCGCGAGTCTCTCACA 87
 DB 36 GAATGACGCGCGCTCTTAGAGAACATATGAAAAACACCGCGAGTCTCTCACA 95
 QY 88 AGCTTAAGTGTCTCTGAGCTGAAGATGACAGGTTTAAAGCCCTGTTCTTTCC 147
 DB 96 AGCTTAAGTGTCTCTGAGCTGAAGATGACAGGTTTAAAGCCCTGTTCTTTCC 155
 QY 148 GTTGTTAATCTATGTTCTTTGGAATTAATAACCTCCCAAGTAGTACTGTTTAA 207
 DB 156 GTTGTTAATCTATGTTCTTTGGAATTAATAACCTCCCAAGTAGTACTGTTTAA 215
 QY 208 TGCTCAACATGCTTTGACCTGTTGAAAAGACCTTTGGCACACATTCGAGGATGCTGAT 267
 DB 216 TGCTCAACATGCTTTGACCTGTTGAAAAGACCTTTGGCACACATTCGAGGATGCTGAT 275
 QY 268 GGAGATGCCATCCATGGAATCAGTGGCGCGCACAGCTATGTTGGT 310
 DB 276 GGAGATGCCATCCATGGAATCAGTGGCGCACAGCTATGTTGGT 318

RESULT 5
 AL135937 59231 bp DNA linear PRI 15-MAR-2001
 LOCUS Human DNA sequence from clone RPL-278022 on chromosome 20. Contains
 DEFINITION STS and GSSs, complete sequence.
 ACCESSION AL135937
 VERSION AL135937.22 GI:7248202
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 59231)
 AUTHORS Phillimore, B.

TITLE
 JOURNAL

COMMENT

Direct Submission
 Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Mar 15, 2000 this sequence version replaced gi:7242336.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr20
 IMPORTANT: This sequence is not the entire insert of clone RPL-278022. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
 The true left end of clone RP4-69788 is at 59132 in this sequence. The true right end of clone RP5-1099D15 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RPL-278022 is from the library RPL-1 constructed by the group of Pieter de Jong. For further details see
 http://www.chori.org/decapac/home.htm
 VECTOR: pCYPAC2

FEATURES
 source
 Location/Qualifiers
 1..59231
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="RP1-278022"
 /clone_id="RPL-1"
 58..213
 /note="MER5A repeat: matches 9. .188 of consensus"
 233..409
 /note="MER3 repeat: matches 6. .191 of consensus"
 572..795
 /note="MIR repeat: matches 3. .241 of consensus"
 2927..2992
 /note="Alu repeat: matches 242. .307 of consensus"
 3019..3635
 /note="LINE2 repeat: matches 5541. .6154 of consensus"
 3880..4411
 /note="LINE4 repeat: matches 4015. .4580 of consensus"
 4412..4600
 /note="MER58A repeat: matches 32. .224 of consensus"
 complement(5199..5709)
 /note="match: GSS: Em:A0613263"
 7115..7585
 /note="Charleia repeat: matches 25. .507 of consensus"
 7631..7695
 /note="MIR repeat: matches 13. .81 of consensus"
 8546..9223
 /note="LINE4 repeat: matches 7117. .7739 of consensus"
 9986..10209
 /note="AluSG repeat: matches 1. .225 of consensus"
 10225..10250
 /note="13 copies 2 mer aa 968 conserved"
 complement(10684..11081)
 /note="match: GSS: Em:A062878"
 11093..11168
 /note="38 copies 2 mer aa 728 conserved"

```

repeat_region 11575..12292
/note="LIM49 repeat: matches 5545..6278 of consensus"
misc_feature 12860..13177
/note="match: GSS: Em:AQ035618"
repeat_region 13327..13423
/note="L2 repeat: matches 2657..2750 of consensus"
repeat_region 13719..13813
/note="L2 repeat: matches 2195..2293 of consensus"
repeat_region 13814..14087
/note="LIM45 repeat: matches 6025..6298 of consensus"
repeat_region 14088..14219
/note="L2 repeat: matches 2078..2195 of consensus"
repeat_region 14965..15067
/note="MIR repeat: matches 169..262 of consensus"
misc_feature complement(15295..15803)
/note="match: GSS: Em:AQ754141"
15849..16248
/note="match: GSS: Em:AQ694331"
15961..16498
/note="LIME3A repeat: matches 5427..5964 of consensus"
16499..16808
/note="AluSx repeat: matches 1..310 of consensus"
repeat_region 16809..17004
/note="LIME3A repeat: matches 5964..6159 of consensus"
17356..17544
/note="MIR repeat: matches 26..230 of consensus"
18267..18357
/note="MIR repeat: matches 50..139 of consensus"
18358..18690
/note="MER58B repeat: matches 2..341 of consensus"
18691..18717
/note="MIR repeat: matches 24..50 of consensus"
18885..18952
/note="34 copies 2 mer ga 79% conserved"
complement(19364..19600)
/note="match: GSS: Em:AQ035738"
22132..22242
/note="LIR33 repeat: matches 410..521 of consensus"
22243..22551
/note="AluSx repeat: matches 1..310 of consensus"
22552..22951
/note="LIR33 repeat: matches 17..410 of consensus"
23356..23643
/note="AluSg repeat: matches 1..288 of consensus"
24124..24510
/note="LIR16C repeat: matches 1..367 of consensus"
24747..24810
/note="32 copies 2 mer ta 68% conserved"
24967..25054
/note="LIM1 repeat: matches 5678..5763 of consensus"
25134..25443
/note="AluSg repeat: matches 1..313 of consensus"
25767..26069
/note="AluSg repeat: matches 1..304 of consensus"
26295..26579
/note="LIMC/D repeat: matches 5290..5579 of consensus"
27383..27481
/note="MIR repeat: matches 20..142 of consensus"
28338..28469
/note="LIMC4 repeat: matches 7841..7973 of consensus"
28472..28583
/note="56 copies 2 mer ta 73% conserved"
28584..28970
/note="LIMC4 repeat: matches 7430..7849 of consensus"
30290..30876
/note="L2 repeat: matches 2124..2658 of consensus"
30877..31223
/note="MER1B repeat: matches 1..337 of consensus"
31224..31421
/note="L2 repeat: matches 1945..2124 of consensus"
31422..31739
/note="AluY repeat: matches 1..309 of consensus"
31750..32046

```

```

repeat_region /note="AluSg repeat: matches 3..298 of consensus"
32047..32188
/note="L2 repeat: matches 1809..1937 of consensus"
repeat_region 32189..32474
/note="AluJo repeat: matches 1..293 of consensus"
repeat_region 32475..32713
/note="L2 repeat: matches 1420..1809 of consensus"
repeat_region 34869..34938
/note="MIR repeat: matches 191..261 of consensus"
repeat_region 35160..35351
/note="MER5A repeat: matches 7..189 of consensus"
repeat_region 36121..36273
/note="MIR repeat: matches 66..225 of consensus"
repeat_region 37904..37941
/note="19 copies 2 mer aa 81% conserved"
repeat_region 38788..38840
/note="MIR repeat: matches 37..84 of consensus"
repeat_region 38841..39254
/note="MSRB repeat: matches 1..426 of consensus"
39255..39304
/note="MIR repeat: matches 84..150 of consensus"
repeat_region 39310..39339
/note="LIM3D repeat: matches 251..280 of consensus"
repeat_region 39314..39717
/note="LIM3E repeat: matches -879..779 of consensus"
39724..39990
/note="LIM49 repeat: matches 3735..4011 of consensus"
39992..40126
/note="AluJo repeat: matches 2..134 of consensus"
40153..40379
/note="LIM49 repeat: matches 3999..4226 of consensus"
40380..40675
/note="AluJo repeat: matches 1..306 of consensus"
40676..42726
/note="LIM49 repeat: matches 4226..6308 of consensus"
43333..43524
/note="LIR16C repeat: matches 165..386 of consensus"
complement(43484..44004)
/note="match: GSS: Em:AQ800109"
44765..44818
/note="MIR repeat: matches 200..250 of consensus"
complement(45503..46337)
misc_feature

```

Query Match 31.5%; Score 119; DB 9; Length 59231;
 Best Local Similarity 100.0%; Pred. No. 5.3e-25;
 Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 148 GTGCTTAATCTAATGCTTCTTGGAATTAATAAACCCTCGCCAGTACTGCTTTTA 207
|||||
Db 18738 GTGCTTAATCTAATGCTTCTTGGAATTAATAAACCCTCGCCAGTACTGCTTTTA 18797
|||||
QY 208 TGCTCAACATGCTTGAAGAGACCTTTGGCACACATTAAGGATGCTGA 266
|||||
Db 18798 TGCTCAACATGCTTGAAGAGACCTTTGGCACACATTAAGGATGCTGA 18856
|||||

```

RESULT 6
 AC119775 162394 bp DNA linear HTG 18-JUN-2002
 LOCUS Rattus norvegicus clone CH230-464J18, *** SEQUENCING IN PROGRESS
 DEFINITION *** 61 unordered pieces.

ACCESSION AC119775
 VERSION AC119775.4 GI:21747184
 KEYWORDS HTG; HTGS; PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 162394)
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barberia,J., Benton,J., Bimoge,K., Blankenburg,K., Bonnin,D.,

* 57055 59393: contig of 2339 bp in length
 * 59394 59493: gap of unknown length
 * 59494 61754: contig of 2261 bp in length
 * 61755 61854: gap of unknown length
 * 61855 64055: contig of 2201 bp in length
 * 64056 64155: gap of unknown length
 * 64156 67129: contig of 2974 bp in length
 * 67130 67229: gap of unknown length
 * 67230 69622: contig of 2393 bp in length
 * 69623 69722: gap of unknown length
 * 69723 72435: contig of 2713 bp in length
 * 72436 72535: gap of unknown length
 * 72536 75742: contig of 3207 bp in length
 * 75743 75843: gap of unknown length
 * 75844 77891: contig of 2049 bp in length
 * 77892 77991: gap of unknown length
 * 77992 80486: contig of 2495 bp in length
 * 80487 80586: gap of unknown length
 * 80587 83986: contig of 3400 bp in length
 * 83987 84087: gap of unknown length
 * 84088 86359: contig of 2273 bp in length
 * 86360 86459: gap of unknown length
 * 86460 91249: contig of 4790 bp in length
 * 91250 91349: gap of unknown length
 * 91350 94806: contig of 3457 bp in length
 * 94807 94906: gap of unknown length
 * 94907 98994: contig of 4088 bp in length
 * 98995 99094: gap of unknown length
 * 99095 105047: contig of 5953 bp in length
 * 105048 105147: gap of unknown length
 * 105148 108668: contig of 3521 bp in length
 * 108669 108768: gap of unknown length
 * 108769 111829: contig of 3061 bp in length
 * 111830 111929: gap of unknown length
 * 111930 116397: contig of 4468 bp in length
 * 116398 116497: gap of unknown length
 * 116498 122101: contig of 5604 bp in length
 * 122102 122201: gap of unknown length
 * 122202 126352: contig of 4151 bp in length

Query Match Best Local Similarity 73.1%; Score 67.8; DB 2; Length 162394; Pred. No. 1.7e-09;

Matches 87; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 148 GTTGTATATCTAATGTTCTTGAATATAAAACCTCCCTGCCAGTACTGTTTGA 207
 Db 79083 GTGTTTATCTAATAAGTTCCCTGAATATAAAACCTCCCTGCACTGAAGACTTGATTA 79024
 QY 208 TGCACACATGCTTGTGCTGTAAGAGACCTTTGGACACATGTAAGGATGGTGA 266
 Db 79023 ATCTCAACATGCTTGTGCTGTAAGAGACCTTTGGACACATGTAAGGATGGTGA 78965

RESULT 7
 LOCUS AL713981 223604 bp DNA linear ROD 28-JUN-2002
 DEFINITION Mouse DNA sequence from clone Rp23-173012 on chromosome 2, complete
 sequence.
 ACCESSION AL713981
 VERSION AL713981.16 GI:21627973
 KEYWORDS HTG.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 223604)

REFERENCE
 AUTHORS Dunn, M.
 TITLE Direct Submission
 JOURNAL Submitted (27-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequests@sanger.ac.uk
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Jun 28, 2002 this sequence version replaced gi:21531387.
 COMMENT ----- Genome Center

Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WormPeP; Information on the WormPeP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-173012 is from the RPI-23 Mouse PAC Library constructed by the group of Peter de Jong. For further details see <http://www.choil.org/bacpac/home.htm> VECTOR: pBAC3.0.

FEATURES
 source
 Location/Qualifiers
 1..223604
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="RP23-173012"
 /clone_id="RPI-23"
 BASE COUNT 62140 a 46575 c 46966 g 67923 t

BASE COUNT 62140 a 46575 c 46966 g 67923 t
 ORIGIN

Query Match Best Local Similarity 71.4%; Score 53.8; DB 10; Length 223604; Pred. No. 2.9e-05;

Matches 85; Conservative 0; Mismatches 32; Indels 2; Gaps 1;

QY 148 GTTGTATATCTAATGTTCTTGAATATAAAACCTCCCTGCCAGTACTGTTTGA 207
 Db 178805 GTGTTTATCTAATAAGTTCCCTGAATATAAAACCTCCCTGCACTGAAGACTTGATTA 178864
 QY 208 TGCACACATGCTTGTGCTGTAAGAGACCTTTGGACACATGTAAGGATGGTGA 266
 Db 178865 ATCTCAACATGCTTGTGCTGTAAGAGACCTTTGGACACATGTAAGGATGGTGA 178921

RESULT 8
 LOCUS AL845312 176543 bp DNA linear HTG 09-AUG-2002
 DEFINITION Danio rerio clone DKEY-287F10, *** SEQUENCING IN PROGRESS ***
 unordered pieces.
 ACCESSION AL845312
 VERSION AL845312.1 GI:22205007
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE zebrafish.
 ORGANISM Danio rerio.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 176543)

REFERENCE
 AUTHORS Burton, J.
 TITLE Direct Submission
 JOURNAL Submitted (04-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
 Clone requests: clonerequests@sanger.ac.uk
 ----- Genome Center
 COMMENT
 Center code: SC
 Web site: <http://www.sanger.ac.uk>

Contact: zface@anger.ac.uk

----- Project Information
Center project name: zk287f10

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 175506 bases at least Q40

Consensus quality: 175799 bases at least Q30

Consensus quality: 176033 bases at least Q20

Insert size: 176343; sum-of-contrigs

Insert size: 177563; 4.7% error; agarose-fp

Quality coverage: 10.13x in Q20 bases; sum-of-contrigs quality

coverage: 10.57x in Q20 bases; agarose-fp

----- NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 70191: contig of 70191 bp in length
* 70192 70291: gap of 100 bp
* 70292 166519: contig of 96228 bp in length
* 166520 166619: gap of 100 bp
* 166620 176543: contig of 9924 bp in length.

FEATURES

source

1. 176543

/organism="Danio rerio"

/db_xref="taxon:7955"

/clone.lib="DKEY-287F10"

/clone.lib="Damokey"

1. 70191

/note="assembly-fragment:03314"

fragment.chain:1"

70292. 166519

/note="assembly-fragment:02085"

fragment.chain:1"

166620. 176543

/note="assembly-fragment:01906.0"

BASE COUNT 57205 a 31955 c 31568 g 55615 t 200 others

ORIGIN

Query Match 10.8%; Score 41; DB 2; Length 176543;

Best Local Similarity 56.2%; Pred. No. 0.21;

Matches 77; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

OY 200 TGGTTTATGCTCAACATGCTTGGACCTTGGAAAAGACCTTTGGCAGACATTTGAAGGG 259

DB 20135 TGGACGCTGCTCAACATGCAATTAATGTGATGAACCAATGACATTTTGAAGGG 20076

OY 260 ATGTGATGAGATGCCAATCCATGCAATGAGTGGGCGGACCTATGTTGCTATAGC 319

DB 20075 ATGTTTCAGATTAATAAATAATAGTCGATCACAAGGCTTCCAAAGTTTTCATGCAATTTTC 20016

OY 320 AGAAGCTCTCTGGCAA 336

DB 20015 GCGTGCTTCATTTTAA 19999

RESULT 9

AC129650 176030 bp DNA linear HTG 31-JUL-2002

LOCUS AC129650

DEFINITION Rattus norvegicus clone CH230-9H1, *** SEQUENCING IN PROGRESS ***,

67 unordered pieces.

AC129650

VERSION AC129650.1 GI:22024419

KEYWORDS HTG; HTGS_PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE

AUTHORS

1 (bases 1 to 176030)

Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Albrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T.,

Barberia,J., Benton,J., Blinze,K., Blankenburg,K., Bonnin,D.,

Bouck,J., Boyle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,

Bunay,C., Butch,P., Burkelt,C., Burrell,K.L., Byrd,N.C.,

Carroll,T.F., Carter,M., Cavazos,S.R., Chachyo,U., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Dayla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,

Douthwaite,K.J., Draper,H., Dugan-Kocha,S., Durbin,K.J.,

Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,

Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,

Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Hawley,P., Hawes,A., Hernandez,J.,

Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,

Homsl,E., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E.,

Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudan,S.,

Karlsson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,

Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

Li,J., Li,Z., Lichtarge,O., Llew.C., Liu,J., Liu,W., Louised,H.,

Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

Maheshwari,M., Mapua,P., Marlin,R., Martindale,A., Martinez,E.,

Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,

Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,

Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,

Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogum,M., Okuwon,G.,

Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,

Peters,J., Pickens,R., Primus,E., Pul,L.L., Quiles,M., Ren,Y.,

Rivers,M., Rojas,A., RojudoKan,I., Roite,M., Ruiz,S., Savery,G.,

Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,

Sutongren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,

Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,

Taney,J., Taylor,C., Taylor,T., Taliford,B., Thomas,N., Thomas,S.,

Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,

Williams,G., Williams,A., Wleczky,R., Woodson,S., Worley,K.,

Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

Weinstock,G., and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 176030)

Worley,K.C.

Direct Submission

Submitted (31-JUL-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GDFP

Center clone name: CH230-9H1

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 115101 bases at least Q40

Consensus quality: 120951 bases at least Q30

Consensus quality: 125022 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 67 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074

REFERENCE
 2 (bases 1 to 152822)
 AUTHORS Abbott,S., Cotton,M. and Spalding,L.
 TITLE The sequence of Homo sapiens BAC clone RP11-44D21
 JOURNAL Unpublished (2001)

REFERENCE
 3 (bases 1 to 152822)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (31-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE
 4 (bases 1 to 152822)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAR-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE
 5 (bases 1 to 152822)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (30-MAR-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE
 6 (bases 1 to 152822)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (20-APR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA On Mar 30, 2002 this sequence version replaced g1.19570183.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: saplens@watson.wustl.edu
 ----- Summary Statistics

 Center project name: H_NH0044D21

COMMENT

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RPc1-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Moon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catalanese,J.J., and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
 Actual start of this clone is at base position 1 of RP11-44D21;
 actual end is at base position 152822 of RP11-44D21.

FEATURES	source	Location/Qualifiers
repeat_region		1..152822
repeat_region		/organism="Homo sapiens"
repeat_region		/db_xref="Caxon:9606"
repeat_region		/chromosome="4"
repeat_region		/map="4"
repeat_region		/clone="RP11-44D21"
repeat_region		/clone_11b="RPc1-11"
repeat_region		6..98
repeat_region		/rpt_family="MaLR"
repeat_region		243..393
repeat_region		/rpt_family="MIR"
repeat_region		1632..1748
repeat_region		/rpt_family="L1"
repeat_region		1747..1819
repeat_region		/rpt_family="L1"
repeat_region		1933..2538
repeat_region		/rpt_family="L1"
repeat_region		2544..2608
repeat_region		/rpt_family="ERV1"
repeat_region		2633..2993
repeat_region		/rpt_family="L1"
repeat_region		3286..3445
repeat_region		/rpt_family="MIR"
repeat_region		5195..5225
repeat_region		/rpt_family="(?)n"
repeat_region		5315..5396
repeat_region		/rpt_family="L2"
repeat_region		6758..7060
repeat_region		/rpt_family="Alu"
repeat_region		8193..8349
repeat_region		/rpt_family="MERL_type"
repeat_region		8366..8496
repeat_region		/rpt_family="L2"
repeat_region		9364..9454
repeat_region		/rpt_family="MIR"
repeat_region		10509..10532
repeat_region		/rpt_family="AT_rich"
repeat_region		15923..16255
repeat_region		/rpt_family="Alu"
repeat_region		20537..20557
repeat_region		/rpt_family="AT_rich"
repeat_region		21092..21123
repeat_region		/rpt_family="AT_rich"
repeat_region		22318..22496
repeat_region		/rpt_family="MIR"
repeat_region		22629..22661
repeat_region		/rpt_family="L2"
repeat_region		22771..22936
repeat_region		/rpt_family="MaLR"
repeat_region		22991..23068
repeat_region		/rpt_family="MaLR"
repeat_region		23202..23569
repeat_region		/rpt_family="L2"
repeat_region		23750..23822
repeat_region		/rpt_family="MaLR"
repeat_region		23823..24562
repeat_region		/rpt_family="L1"
repeat_region		24563..24965
repeat_region		/rpt_family="MaLR"
repeat_region		26011..26226
repeat_region		/rpt_family="MIR"
repeat_region		26418..26454
repeat_region		/rpt_family="AT_rich"
repeat_region		26710..27038
repeat_region		/rpt_family="MERL_type"
repeat_region		28399..28695
repeat_region		/rpt_family="Alu"
repeat_region		29527..29789
repeat_region		/rpt_family="Alu"
repeat_region		30235..30280
repeat_region		/rpt_family="CR1"
repeat_region		31674..32256

```

repeat_region      /rpt_family="ERV1"
                    32257..33296
                    /rpt_family="ERV1"
repeat_region      36621..36659
                    /rpt_family="(T)n"
repeat_region      36842..37055
                    /rpt_family="L1"
repeat_region      37056..37334
                    /rpt_family="L1"
repeat_region      37408..37792
                    /rpt_family="L1"
repeat_region      37872..38039
                    /rpt_family="L1"
repeat_region      38041..38088
                    /rpt_family="(TACA)n"
repeat_region      38156..38390
                    /rpt_family="ERV1"
repeat_region      38399..38987
                    /rpt_family="ERV1"
repeat_region      38987..39739
                    /rpt_family="ERV1"
repeat_region      39740..39929
                    /rpt_family="ERV1"
repeat_region      39930..40066
                    /rpt_family="ERV1"
repeat_region      40159..41190
                    /rpt_family="L1"
repeat_region      41238..42323
                    /rpt_family="L2"
repeat_region      42632..42829
                    /rpt_family="L2"
repeat_region      43009..43674
                    /rpt_family="L2"
repeat_region      43675..44371
                    /rpt_family="L2"
repeat_region      44375..44447
                    /rpt_family="L2"
repeat_region      44898..44934
                    /rpt_family="(TG)n"
repeat_region      45261..45307
                    /rpt_family="(TC)n"
repeat_region      45308..45339
                    /rpt_family="(CA)n"

```

```

Query Match      10.2% Score 38.4; DB 9; Length 152822;
Best Local Similarity 58.9% Pred. No. 1.3;
Matches 66; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```

```

OY 92 TGAATGTGTCTGAGCTGAGAGATGACGCTTTAAGCCCTGTTCTTCCGTTG 151
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 120753 TGAATGATTAATCTCTAGTAGAGAGCTGCTGTTTACTATCCCTTTTCTGTAA 120812
    ||| ||||| | | | | | | | | | | | | | | | | | | | | | | |
OY 152 TTTAATCTAATGTTCTTGGAAATAAACCTCCCTCCAGATGACTTGT 203
    ||| ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB 120813 CTTACTCTAATAATGCTTTGAAAAATACATGCTTAATAATCATATTAGT 120864
    ||| ||||| | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 13
AC080170/c 190912 bp DNA linear HTG 24-AUG-2002
LOCUS      Homo sapiens chromosome 10 clone RP11-435B15 map 10, WORKING DRAFT
DEFINITION
ACCESSION  AC080170
VERSION    AC080170.2 GI:12830207
KEYWORDS   HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 190912)
            Birren,B., Linton,L., Nusbaum,C. and Lander,E.
            TITLE      Homo sapiens chromosome 10, clone RP11-435B15
            JOURNAL
            Unpublished

```

REFERENCE

AUTHORS

2 (bases 1 to 190912)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L.,
 Boukhalter,B., Brown,A., Burkett,G., Campolongo,A., Castle,A.,
 Choepel,Y., Collangelo,M., Collins,S., Collamore,A., Cooke,P.,
 Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
 Fitzhugh,W., Gage,D., Galagan,J., Gargyala,S., Ginde,S., Goyette,M.,
 Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
 Iliev,I., Johnson,R., Jones,C., Kamm,L., Karatas,A., Larocque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,K., Lieu,C., Liu,G.,
 Macdonald,P., Margulis,N., McCarthy,M., McEwan,P., McKernan,K.,
 Mcpheeters,R., Meldrim,J., Menus,L., Mihova,T., Mlenga,V.,
 Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
 O'Donnell,P., O'Neill,D., Pollara,V., Raymond,C., Riedack,M., Riley,R.,
 Piere,N., Pisanl,C., Pollara,V., Raymond,C., Riedack,M., Riley,R.,
 Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
 Sougnez,C., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
 Strausz,N., Subramanian,A., Talamas,J., Teafaye,S., Theodore,J.,
 Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zalnoun,J.,
 Zimmer,A. and Zody,M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (28-SEP-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 190912)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
 Camarata,J., Campolongo,A., Choepel,Y., Collangelo,M., Collins,S.,
 Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
 Gargyala,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Karatas,A., Larocque,K., Lamazares,R., Landers,T.,
 Lehoczy,J., Levine,K., Liu,G., Maclean,C., McDonald,P.,
 Margulis,N., Mathews,C., McCarthy,M., McEwan,P., McKernan,K.,
 Mcpheeters,R., Meldrim,J., Menus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunhkhang,P., Piere,N., Pollara,V., Raymond,C., Riedack,M.,
 Riedack,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
 Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
 Sougnez,C., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
 Strausz,N., Subramanian,A., Talamas,J., Teafaye,S., Theodore,J.,
 Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zalnoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

COMMENT

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 14, 2001 this sequence version replaced gi:10334889.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L10588
 Center clone name: 435_B_15

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 184508 bases at least Q40
 Consensus quality: 187307 bases at least Q30
 Consensus quality: 188400 bases at least Q20
 Insert size: 198000; agarose-ftp
 Insert size: 189212; sum-of-ctrls
 Quality coverage: 4.9 in Q20 bases; agarose-ftp
 Quality coverage: 5.1 in Q20 bases; sum-of-ctrls

1	8804:	contig of 8804 bp	in length
8805	8904:	gap of 100 bp	in length
8805	9931:	contig of 1087 bp	in length
9992	10091:	gap of 100 bp	in length
10092	11974:	contig of 1883 bp	in length
11975	12074:	gap of 100 bp	in length
12075	14089:	contig of 2015 bp	in length
14090	14189:	gap of 100 bp	in length
14190	15697:	contig of 1508 bp	in length
15698	15797:	gap of 100 bp	in length
15798	18681:	contig of 3064 bp	in length
18682	18961:	gap of 100 bp	in length
18962	21671:	contig of 2710 bp	in length
21672	21771:	gap of 100 bp	in length
21772	23839:	contig of 2068 bp	in length
23840	23939:	gap of 100 bp	in length
23940	28763:	contig of 4844 bp	in length
28784	28883:	gap of 100 bp	in length
28884	33590:	contig of 4707 bp	in length
33591	33690:	gap of 100 bp	in length
33691	42274:	contig of 8584 bp	in length
42275	42374:	gap of 100 bp	in length
42375	5118:	contig of 8744 bp	in length
5119	51218:	gap of 100 bp	in length
51219	63841:	contig of 12623 bp	in length
63842	63941:	gap of 100 bp	in length
63942	76817:	contig of 12876 bp	in length
76818	76917:	gap of 100 bp	in length
76918	90127:	contig of 13210 bp	in length
90128	90227:	gap of 100 bp	in length
90228	136743:	contig of 46516 bp	in length
136744	136943:	gap of 100 bp	in length
136944	173143:	contig of 36300 bp	in length
173144	173243:	gap of 100 bp	in length
173244	190912:	contig of 17669 bp	in length

misc_feature	/note="assembly-fragment"	33691..42274
misc_feature	/note="assembly-fragment"	42375..51118
misc_feature	/note="assembly-fragment"	51219..63841
misc_feature	/note="assembly-fragment"	63942..76817
misc_feature	/note="assembly-fragment"	76918..90127
misc_feature	/note="assembly-fragment"	90228..136743
misc_feature	/note="assembly-fragment"	136844..173143
misc_feature	/note="assembly-fragment"	173244..190912
misc_feature	/note="assembly-fragment"	clone_end:T7 vector_side:right"
BASE COUNT	57903 a 33982 c 34268 g 63050 t 1709 others	
ORIGIN		
Query Match	10.2% ; Score 38.4; DB 2; Length 190912;	
Best Local Similarity	58.9% ; Pred. No.1.3;	
Matches 66; Conservative	0; Mismatches 46; Indels 0; Gaps 0;	
QY 92 TGAATGCTGTTCGGAGCTCAGAAGTACACCGTTGTTAAAGCCCTGTCTTCCGCTG 151		
Db 136670 TGAATGATTAAATTCCTAGTGTTGGTAGACCCTGGTTTAAGTATCCCTGTTTTCTGTAA 136611		
OY 152 TTATATCATATGCTCTTGGAAATAAACAACCTCCCTGCCCAAGTACTTGTCT 203		
Db 136610 CTTACTCMAAATGTTTGGAAAAATATACATGTCATATATCATATTACT 136559		
RESULT 14		
AC024785/c	30960 bp DNA linear INV 24-MAY-2002	
LOCUS	Caeenorhabditis elegans cosmid Y46C8AL, complete sequence.	
ACCESSION	AC024785	
VERSION	AC024785.1 GI:7140340	
KEYWORDS	HTG.	
SOURCE	Caeenorhabditis elegans.	
ORGANISM	Caeenorhabditis elegans	
REFERENCE	Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitidae; Rhabdilitidae; Peloderinae; Caeenorhabditis. 1 (bases 1 to 30960)	
AUTHORS	Waterston,R.	
TITLE	Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium Science 282 (5396), 2012-2018 (1998)	
JOURNAL	99069613	
MEDLINE	9851916	
PUBMED	2 (bases 1 to 30960)	
REFERENCE	Cordes,M.	
AUTHORS	The sequence of C. elegans cosmid Y46C8AL	
TITLE	Unpublished (2001)	
JOURNAL	3 (bases 1 to 30960)	
REFERENCE	Waterston,R.H.	
AUTHORS	Direct Submission	
TITLE	Submitted (01-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA	
JOURNAL	4 (bases 1 to 30960)	
REFERENCE	Waterston,R.	
AUTHORS	Direct Submission	
TITLE	Submitted (28-MAR-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	
JOURNAL	5 (bases 1 to 30960)	
REFERENCE	Waterston,R.	
AUTHORS	Direct Submission	
TITLE	Submitted (28-JUN-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA	

CDS

88-Sequence"
 join(16234..16362,17175..17391,17491..17739,18508..19286,
 19338..19628,20472..20616,20666..20763,20809..20930,
 20975..21008)
 /gene="Y46C8AL.6"
 /product="Hypothetical protein Y46C8AL.6"
 /protein_id="AAE60603.1"
 /db_xref="GI:7331915"

/translation="MRASBPBAGRPADARRCVMEKCEKRLLEADTFEGVPLSL
 LKSLSTLILGLVSVCPDRNDLEVGLETPVAQMTNDARDCHYKNPLGSSPL
 TYPDSVTSNNLASHARIAPRAKNWISLKRSSGSLSPNGSLPVYVTEGKNG
 KSYFEQVSNNAKMTLDNDTNEFVCSIDSGNPTPELRLQSSVPWISCHDPE
 TLDPHFGEYIEHRESKIPRNNDLTQHDRIHCOLDRLLSLKLNKLISMLAEFLN
 RYFQVKVGDLSSEFRKADCGVAGVSLPLFGFVEVIEINILPEAIKCOFADDLK
 YTAIPSSNNVHLQKAIETVEISKATKLANDKVCISLIGNTTEPQYTIENSIL
 TRSNVNDLGFQIEPRLNFSFMWKKSVKSKFMHOLFNSVCSKRLHLTLXYTFVR

Query Match 10.1%; Score 38.2; DB 3; Length 30960;
 Best Local Similarity 63.7%; Pred. No. 1.5;
 Matches 58; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY 155 AATCAATGCTTGTGGAATAAAACCTCCGCAAGTAGTACTGTTTATGCTCAA 214
 DB 1728 AATCCAGCTGTACGTTAAATATATCTTACATGCCAAGTAGTCTGTTTATGCTCAG 1669

OY 215 CATGCTTGTACTGTGTAAGAAAGACCTTTGG 245
 DB 1668 GAAACATATGCCAGATAGGAAAGACCTACTGG 1638

RESULT 15
 LOCUS AL732628 174196 bp DNA linear HTG 09-AUG-2002

DEFINITION Danio rerio clone CH211-144F18, ** SEQUENCING IN PROGRESS ***, 3
 unordered pieces.

ACCESSION AL732628 GI:22204545
 VERSION AL732628.8

KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE zebrafish.

ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 174196)

REFERENCE Direct Submission
 Submitted (31-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfaces@sanger.ac.uk

Clone requests: clonerequests@sanger.ac.uk
 On Aug. 11, 2002 this sequence version replaced gi:21911655.

COMMENT ***** Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zfaces@sanger.ac.uk
 Project Information
 Center project name: zc144F18

Summary Statistics

Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 173595 bases at least Q40
 Consensus quality: 173741 bases at least Q30
 Consensus quality: 173854 bases at least Q20
 Insert size: 173996; sum-of-contigs
 Insert size: 161892; 19.9% error; agarose-fp
 Quality coverage: 6.72x in Q20 bases; sum-of-contigs Quality
 coverage: 7.22x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 53761: contig of 53761 bp in length
 * 53762 53861: gap of 100 bp
 * 53862 77324: contig of 23463 bp in length
 * 77325 77424: gap of 100 bp
 * 77425 174196: contig of 96772 bp in length.
 Location/Qualifiers

FEATURES

source 1. 174196
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone="CH211-144F18"
 /clone_id="CHOR1-211"
 1. 53761
 /note="assembly_fragment:00869
 fragment_chain:1"
 misc_feature 53862..77324
 /note="assembly_fragment:01503
 fragment_chain:1"
 misc_feature 77425..174196
 /note="assembly_fragment:00262
 fragment_chain:1"

BASE COUNT 59260 a 30996 c 29354 g 54386 t 200 others
 ORIGIN

Query Match 10.0%; Score 37.8; DB 2; Length 174196;
 Best Local Similarity 54.7%; Pred. No. 2;
 Matches 75; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY 200 TGGTTTATGCTCAACATGCTTTGACTGTGTAAGAAAGACCTTGGCACACATTGAAAGG 259
 DB 98366 TGGACGTGTGTTCAACATGCTTCAATATATGTGATAGACCAAGAAAGACCTTTGAAAGG 98307

OY 260 ATGTGTATGAGATATCCATATCATATGATATGATGAGTGGCCAGCTATGTGTATATGAC 319
 DB 98306 AAGTTCAGATATTAACATATGATGATCCCAAGGCTCTCCAAAGTTTCAATGCAATTC 98247

OY 320 ACAAGTCTCTTGGCAA 336
 DB 98246 GGGTGTATTATTATTAA 98230

Search completed: January 31, 2003, 03:21:40
 Job time : 3850 secs